JS Journal *of* Historical Syntax
Volume 8, Article 6: 1–49, 2024 · DOI: 10.18148/hs/2024.v8i6.219

NUMERAL CLASSIFIERS IN YUCATEC MAYA: MICROVARIATION AND SYNTACTIC CHANGE *

Barbara Blaha Pfeiler
Universidad Nacional Autónoma de México, CEPHCIS

Stavros Skopeteas
University of Göttingen

Abstract Numeral classifiers in Yucatec Maya are subject to two processes of language change that create variation in the contemporary language. The first process is the use of a general classifier instead of specific sortal classifiers. The second process is the use of the general classifier along with mensural classifiers. Our study examines the microvariation of the contemporary language in space and time, based on data from the *Atlas of Yucatec Maya* and draws inferences about the entity of change in diachronic perspective. Our findings show that these processes are partially interconnected, reflecting the emergence of a general marker of *Cardinality* (Krifka 1995, Bale & Coon 2014, Bale, Coon & Arcos 2019). The dispersion of these phenomena in geographical space shows that they only partially overlap, suggesting that the underlying processes may apply independently from each other. Furthermore, the use of the general classifier in expressions of measure does not apply equally to all mensural classifiers. Hence, a further source of variation comes from mensural classifiers: some of them lose their function as classifiers and are only used as measure nouns. Contemporary variation can thus be understood as the cumulative effect of these processes.

* We are particularly grateful to Flor Canché Teh for her work in the transcription of this corpus and to Amedee Colli Colli for her consultation as native speaker. Grateful thanks are due to Christian Lehmann (University of Erfurt), Martín Sobrino Gómez (CEPHCIS), Alexander Voss (Universidad Autónoma de Quintana Roo, Chetumal), three anonymous reviewers and the handling editor for their comments on earlier versions of this study. This article is part of the research project ”Yucatec Maya: Variation in time and space”, funded by the CONAHCYT (Consejo Nacional de Humanidades, Ciencias y Tecnologias; No. 277501; www.conahcyt.mx/) and the DFG (Deutsche Forschungsgemeinschaft; www.dfg.de/). The revision of this article in 2024 was part of the project ”Transformaciones socio-económicas, prácticas comunicativas y gramáticas indígenas: comprender los desarrollos recientes en el maya yucateco” (DFG-UNAM-PAPIIT).

1 Introduction

The investigation of microvariation provides an ideal playground for studying the clustering of grammatical properties in languages that minimally differ (Benincà 1989: 3, Kayne 1996: xi-xii, Brandner 2012: 119). If certain properties of different varieties co-vary, they may be reducible to a more abstract grammatical entity. Conversely, the lack of co-variation is equally informative, indicating that the properties at issue are separate entities in the grammar. Since microvariation deals with historically related varieties, such as dialects, it allows for insights into the interconnectedness of processes in language change.

This reasoning underpins the agenda of the present study, which deals with current developments in numeral classifiers of Yucatec Maya. Some speakers employ the general classifier *p'éel* 'CL.UNIT' instead of specific sortal classifiers, such as *túul* 'CL.ANIM'; see (1).[1]

(1)	*jun-túul*	*k'éek'en*	/	*jum-p'éel*	*k'éek'en*
	one-CL.ANIM	pig		one-CL.UNIT	pig
	'one pig'				

Furthermore, some speakers insert the general classifier in measure expressions although these constructions could be formed with an expression of measure alone (without a difference in meaning); see (2).

(2)	*jun-cháach*	*iib*	/	*jum-p'éel*	*cháach*	*iib*
	one-CL.BUNCH	bean		one-CL.UNIT	CL.BUNCH	bean
	'one bunch of beans'					(Briceño Chel 1992: 72)

The question of the present study is whether these phenomena reflect distinct diachronic processes or just result from a single process, namely the

1 Orthography: examples of contemporary Yucatec Maya are uniformly presented in the local orthography (Briceño Chel & Can Tec 2014), independently of the orthography used in the cited sources; in examples from earlier texts, the orthography of the cited source is preserved. Vowels are distinguished in short <*a*>, long with low tone <*aa*>, long with high tone <*áa*>, and rearticulated <*a'a*>. Consonant symbols correspond to the IPA values to the exception of <*j*> [h], <*x*> [ʃ], and <*ch*> [tʃ]. Apostrophe <'> stands for glottal stop after vowels and for glottalization after consonants (<*p'*>, <*t'*>, <*k'*>, <*ts'*>, <*ch'*>).

Data Sources: (a) text collections: (Lehmann 2023), (Monforte, Dzul & Gutiérrez Bravo 2010), (Skopeteas, Colli Colli, Schellenbach, Brokmann, Fischer & Gálvez Wimmelmann 2020); whenever available, speaker metadata are given in parenthesis: sex, age, monolingual/bilingual, place of residence, and year of recording; (b) elicitation: examples without a bibliographical reference were elicited with a female speaker, 53 years old, bilingual (Yucatec Maya and Spanish), resident of Felipe Carrillo Puerto.

generalization of the classifier *p'éel* 'CL.UNIT'. The hypothesis is that the classifier *p'éel* evolves into a general marker of Cardinality accompanying numeral roots in all environments in which they are used as cardinal. The difference between (1) and (2) lies in the function of classifiers. Sortal classifiers as in (1) do not contribute to the meaning and can be substituted without loss of information. Mensural classifiers as in (2) denote units of measurement and cannot be replaced without altering the meaning (see Section 3.5).

The present study examines data from the *Atlas of Yucatec Maya*, an online resource enabling to observe variation in Space along with additional demographic factors (Blaha Pfeiler, Skopeteas & Verhoeven 2022). Yucatec Maya is spoken on the peninsula of Yucatán (Mexico), with a sizable population of 787,594 speakers in the Mexican states of Yucatán, Quintana Roo, and Campeche (census 2010; INEGI 2011), as well as by 2,869 speakers in the northern part Belize (census 2010; Statistical Institute of Belize 2010: 78). The Yucatecan peninsula is a predominantly level landscape covered by tropical forest (in the central, eastern and southern parts) without significant natural barriers like large rivers, mountains or lakes, that could impede communication and create sharp linguistic/dialectal boundaries. Dialectal variation is manifested as a continuum in geographical space, with the major source of diversity being the East-West axis (Tozzer 1921: 14, Edmonson 1986: 2, Blaha Pfeiler & Hofling 2006).

A important facet of the present language situation is the contact to Spanish, since speakers of Yucatec Maya are a minority within the peninsula (approximately 24%). The major density of indigenous speakers lies in the central and eastern regions of the peninsula (Blaha Pfeiler et al. 2022). Over time, the proportion of Mayan speakers has been decreasing. Based on INEGI (2011), among the age group 85 and older (total population: 22,915), 9% were monolingual (n = 2,113) and 44% were bilingual with Spanish (n = 10,240), while among the age group 10-14 years (382,549 in total), only 0.6% were monolingual (n = 2,348) and 11% were bilingual (n = 44,642). These counts indicate that (a) the proportion of Mayan speakers is decreasing over time and (b) monolingual speakers almost disappear, which leads to a generalized bilingualism within the indigenous speakers' population. The current language situation is an instance of 'diglossia-with-bilingualism' (Pfeiler 2014; in terms of Fishman 1967), which means that speakers use both languages in their daily life (possibly in different situations) and take for granted that their interlocutors are competent in both languages when speaking Yucatec Maya. The effects of this situation are evident in language contact, that is reflected in various phenomena, such as Spanish borrowings (Pool Balam & Le Guen 2015), the emergence of definite articles (Vázquez-Rojas Maldonado, García Fajardo, Gutiérrez-Bravo & Pozas Loyo 2018), the obligatorification of plural marking (Gutiérrez Bravo & Uth 2020), among others.

With this background, the present study examines the current microvariation in the use of classifier constructions. After introducing the basic facts in Section 2 and the relevant assumptions in Section 3, Section 4 outlines the methodological background of the present study. The results are reported in Section 5 and interpreted in Section 6, while Section 7 concludes.

2 Basic facts

Numerals obligatorily combine with numeral classifiers in Yucatec Maya, as in (3). Classifiers can be directly attached to the numeral, as in (3)-(5). Numerals and classifiers are written in a single word in the local orthography, which suggests that they are integrated in the same prosodic word under particular phonological conditions (including weight, see Briceño Chel 1993). The inventories of classifiers in grammatical descriptions are long (e.g., 75 classifiers in Beltrán 1859: 203–208, 171 in Thompson 1970, more than 100 in Miram 1983, and approximately 250 in Bricker, Poot Yah & Dzul de Poot 1998) and contain two different types of elements, as illustrated in the following: sortal and mensural classifiers. Sortal classifiers identify the class of the nominal predicate at issue as in (3); (Croft 1994).

(3) (a) animate

jun-(túul)* *k'éek'en*
one-CL.ANIM pig
'one pig'

(b) plant

jun-(kúul)* *che'*
one-CL.PLANT tree
'one tree'

(c) long object

jun-(ts'íit)* *kib*
one-CL.LONG candle
'one candle'

(d) else

jum-(p'éel)* *tuunich*
one-CL.UNIT stone
'one stone'

The classifier *p'éel* 'CL.UNIT' in (3d) is a special case of sortal classifier, since it is a superordinate concept that can substitute specific sortal classifiers. Its use varies between contemporary dialects: while it is restricted to inanimates in eastern/central varieties, it can be used for all types of entities in western varieties (Martínez Corripio 2003: 85); see examples in (4a-c). Across varieties, *p'éel* 'CL.UNIT' can refer to any countable entity apart from kinds: *ka'a-p'éel wáaj* (two-CL.UNIT tortilla) is necessarily 'two units of tortilla' and not 'two kinds of tortilla'.

(4) (a) *jum-*(p'éel) k'éek'en*
one-CL.UNIT pig
'one pig' (in western varieties)

(b) *jum-*(p'éel) che'*
one-CL.UNIT tree
'one tree'

(c) *jum-*(p'éel) kib*
one-CL.UNIT candle
'one candle'

Beyond the sortal classifiers in (3), Yucatec Maya has various mensural classifiers, which have a different function: within numeral constructions, these elements express the units that are counted by the numerals (Croft 1994); see (5).

(5) (a) *jun-*(luuch) ja'*
one-CL.GOURD water
'one gourd water'

(b) *jun-*(cháach) iib*
one-CL.BUNCH bean
'one bunch of beans' (Briceño Chel 1992: 72)

In contemporary Yucatec Maya, the construction in (5) alternates with a multiple-classifier construction where the general classifier is combined with mensural classifiers, as in (6); cf. (5b) (Briceño Chel 1992: 72, Briceño Chel 1993: 118). This construction is characteristic for mensural classifiers and infrequent but possible with some sortal classifiers (see Section 3.5 and results in Sections 5.1-5.2).[2]

2 The classifier *p'éel* does never occur with the classifier *túul*.

(6) *jum-p'éel* *cháach* *iib*
one-CL.UNIT CL.BUNCH bean
'one bunch of beans.' (Briceño Chel 1992: 72)

Numeral classifiers originate in nouns or deverbal nominalizations. Objects used in measuring are often the source of mensural classifiers, e.g., *leek* 'CL.GOURD' from noun *leek* 'round gourd' (Bricker et al. 1998: 367). Many classifiers are deverbal nouns, derived with a high tone from transitive verb roots (*p'éel* 'CL.UNIT' < *p'el* 'shell', *xéet'* 'CL.PIECE' < *xet'* 'break apart') or positional verb roots (*jáaw* 'CL.GOURD' < *jaw* 'lie face up'). High-tone derivation from verbal roots is also found in nominalizations, e.g., noun *ch'óop* 'blind' < verb *ch'op* 'injure the eyes' (Martin Sobrino Gómez, p.c.).

Numeral classifiers are already attested in Classical Maya (hieroglyphic inscriptions and codices of the Pre-Columbian era), whereby it is disputed whether they were obligatory or optional at this stage; see Macri (2000: 19) for the former and Law & Stuart (2017: 166) for the latter view. In Colonial Yucatec Maya (1500–1750 CE), numeral classifiers were certainly optional (Lehmann 2020: 807); see (7).

(7) *ox* *cul-ic* *c̄hic̄h* *ca* *in* *hul-ah=e*
three sit-DEP bird CONJ A.1SG shoot-CMPL=R3
'three birds were sitting together and I shot them.' (Coronel 1620 [1929]: 46, cited from Lehmann 2020: 807)

The variation in the use of mensural classifiers (with or without general classifier, as in (6) and (5)) is also found in earlier texts; see *Contrata de una maya de yucatan* (1849 CE, Ts'ilam) (Barrera Vázquez 2012): *hun ts'am xanakeuel* (one CL.SET sandal) 'a pair of sandals' [line 102] and *ca-p'eel ts'am nook* (two-CL.UNIT CL.SET cloth) 'two sets of clothes' [line 97]. Multiple classifiers of this type are also reported for further Mayan languages, e.g., Chontal (Keller 1955: 271) and Akatek (Zavala 2000). In Akatek, multiple classifiers are reported to equally appear with sortal and mensural classifiers.

A further diachronic development is the recategorization of classifiers to nouns. For instance, temporal intervals ('day', 'year', '20-year period', etc.) were used as mensural classifiers already in Classical Maya (Macri 2000: 21–25). In Colonial Maya, the word for 'day' is either used as a classifier, e.g. *hun-kin* (one-CL.DAY) 'one day' [p. 254] or in combination with the general classifier, e.g. *hun-ppel kin*[3] (one-CL.UNIT CL.DAY) 'one day' [p. 331] (*Manuscrito Morley*, 1576 CE) (Whalen 2003). In the contemporary language, the use of

3 Orthography of the original: <pp> stands for /p'/.

these words with the general classifier is predominant: for instance, in the texts of Lehmann (2023), all nine instances of the word *kin* 'day' with numerals occur with *p'éel* 'CL.UNIT', which indicates that the time expression *kin* 'day' has been recategorized as a noun.

The generalization of a single classifier is often attributed to language contact. Berlin (1968: 23) observes that speakers with greater exposition to Spanish prefer the construction with the general classifier instead of specific classifiers in Tzeltal, De León Pasquel (1988: 78) discusses the use of general classifier as an instance of pattern borrowing from Spanish. The dominant use of a single classifier is accompanied by the shrinkage of the inventory of numeral classifiers over time, which is also reported to correlate with the speaker's exposition to Spanish (Bricker 2019: 244).

In sum, there are two sources of variation in Yucatec Maya: the substitution of sortal classifiers by a general classifier and the mediation of the general classifier between numerals and mensural classifiers. Historical facts suggest various diachronic developments: (a) the emergence of a general classifier that applies with all nouns, (b) the shrinkage of the inventory of numeral classifiers over time, (c) the recategorization of earlier classifiers as nouns.

3 Deconstructing classifier constructions

3.1 Background

Numeral constructions vary between languages: in languages such as English ("non-generalized classifier languages"), numerals can directly combine with count nouns to form numeral expressions; in languages such a Chinese ("generalized classifier languages"), numerals combine with nouns always with the mediation of a classifier. Both types of languages may have expressions of measure that mediate between the numeral and the nominal projection; but only non-generalized classifier languages have a class of nouns that can be directly used with numerals.

Two paradigms of accounts have been proposed in order to assess the variation in the use of classifiers (see Bale & Coon 2014, Little, Moroney & Royer 2022). According to the 'classifiers for nouns' view, the source of cross-linguistic variation lies in the possibility of nouns to be used with atomic denotations. In non-generalized classifier languages such as English, a subset of nouns (i.e., count nouns) can be valued as atomic without an overt operator (without a classifier), while in generalized classifier languages, individuation must be overtly expressed by a classifier (Chierchia 1998, see also Lucy 1992, Yu 2023, 2024 on Yucatec Maya). This approach is schematically represented in (8) whereby the entity type ε is unvalued for the nominal root and can be valued as atomic either silently as in (8a) or overtly as in (8b) (examples from Krifka 1995: 399).[4]

(8) Classifiers for nouns

(a) non-generalized classifier languages

	Num [	Ø	$N_{\varepsilon:\text{unvalued}}$ $]_{\varepsilon:\text{ATOMIC}}$
English:	*three*		*bears*

(b) generalized classifier languages

	Num [	Clf	$N_{\varepsilon:\text{unvalued}}$ $]_{\varepsilon:\text{ATOMIC}}$
Chinese:	*sān*	*zhī*	*xíong*
	three	CL	bear

According to the ‘classifiers for numerals’ view, the source of cross-linguistic variation lies in the numerals. In order to connect a number with a noun, it is necessary to specify the exact function of the number: merging the number ’2’ and the denotation of ’apple’ can yield ’a set of 2 apples’, ’the second apple in a row of apples’, etc. Hence, the composition of numbers and nouns is mediated by a *measure function* μ, which can be valued as cardinal (indicating the number of atomic entities of a set, e.g., *two*), ordinal (indicating the position in an ordered list, e.g., *second*), distributive (indicating the number of entities of subgroups, e.g., *two by two*), multiplicative (indicating the number of iterations, e.g., *twofold*, *double*), etc. (Krifka 1995, Scontras 2013: 10, Bale & Coon 2014: 697, Scontras 2022: 1184, Zabbal 2005: 3). Numerals denote numbers, but vary cross-linguistically regarding the exponence of the measure function (Krifka 1995, Bale & Coon 2014, Bale et al. 2019): some numerals can be used as cardinals without an overt exponent of Cardinality (i.e., the measure function is silently valued as ‘cardinal’), such as the English numerals; see (9a). Other numerals can form cardinal expressions only in combination with classifiers, which expound the cardinal value of the measure function (among else), as in (9b).

(9) Classifiers for numerals

(a) non-generalized classifier languages

	[$\text{Num}_{\mu:\text{unvalued}}$	$Ø]_{\mu:\text{CARDINAL}}$	N
English:	*three*		*bears*

4 Semantic theories vary regarding the analysis of the denotation of the nominal root; for our purposes, the only relevant issue is whether the valuation as ‘atomic’ must be overtly expressed or not.

(b) generalized classifier languages

	[$\text{Num}_{\mu:\text{unvalued}}$	Clf $]_{\mu:\text{CARDINAL}}$	N
Chinese:	*sān*	*zhī*	*xíong*
	three	CL	bear

The approaches in (8) and (9) differ regarding the relevance of the count-mass distinction. Crucially, the 'classifiers for nouns' approach entails that languages of the type (8a) have a count-mass distinction such that some nouns can be valued as atomic without a classifier, while languages of the type (8b) do not do so (see 'nominal mapping parameter' in Chierchia 1998; see also Lucy 1992: 72–75 on Yucatec Maya). The 'classifiers for numerals' approach in (9) does not rely on this distinction: languages with a count-mass distinction may belong to either type (9a) or type (9b) depending on the properties of their numerals (Bale & Coon 2014: 705).

Little et al. (2022) provide evidence that the analyses in (8b) and (9b) correspond to two subtypes of generalized classifier languages and propose diagnostics for the typological classification. If the classifiers are required by the numerals, then:

- the occurrence of classifiers may be conditioned by numerals (such that only a subset of the numerals requires classifiers);
- classifiers also appear with numerals in the absence of nouns (e.g., in counting).

If the classifiers are required by the nouns, then:

- the occurrence of classifiers may be conditioned by nouns (such that only a subset of the nouns requires classifiers);
- classifiers also appear with nouns in the absence of numerals (e.g., with demonstratives).

After establishing that classifier constructions differ from complex noun phrases in Section 3.2, we examine numeral classifiers in Yucatec Maya in view of these accounts in Section 3.3. Section 3.4 presents further evidence about the functions of numeral roots and Section 3.5 discusses the use of classifiers in expressions of measure. Based on this background, Section 3.6 evaluates the processes of language change introduced in Section 2.

3.2 Classifier constructions differ from complex noun phrases

Numeral-classifier constructions syntactically differ from complex noun phrases. In Yucatec Maya, dependencies between nouns are marked through person agreement on the head noun, as shown by the person marker (A.3) in (10a).[5] With the exception of inalienable nouns, which can appear in possessive constructions without additional marking, head nouns must be relationalized by the suffix *-Vl* '-REL' (whereby the vowel *V* is mostly determined by vowel harmony). Thus, heads of complex noun phrases agree with dependent nominals in person and are relationalized (if necessary), as in (10), which differs from numeral-classifier constructions, in which dependencies are not overtly marked, cf. (3)-(5). Note that nouns used as classifiers may also appear as heads of complex noun phrases, cf. (10b) and (3b) (Lehmann 1998: 63).

(10) (a) *u* *ja'-il* *k'a'anáab*
A.3 water-REL sea
'the water of the sea' (Lehmann 1998: 70)

(b) *k-a* *w-il-ik* (...) *u* *kúul-ul-o'ob* *le* *k'áax-o'ob-o'*
IPFV-A.2 Ø-see-INCMPL[B.3] A.3 plant-REL-PL DEF herb-PL-DIST
'you see plants of herbs' (Skopeteas et al. 2020: YUC-TXT-PA-10)
(SPEAKER: F; 37Y; bilingual; Dzan; rec. 2015)

Summing up, classifiers have nominal properties, but numeral-classifier constructions differ from complex noun phrases. Classifier constructions are formed with simple juxtaposition of the nominal to the classifier, while complex noun phrases contain relational morphology (person agreement and relationalizing affix).

3.3 Classifiers are for numerals

Numeral classifiers are required by the numerals, as already argued by Lehmann (2010) for Yucatec Maya and by Bale & Coon (2014), Bale et al. (2019), Little et al. (2022) for Ch'ol. In terms of the diagnostics in Section 3.1, classifiers are obligatory for certain quantifiers and must appear with these elements even in the absence of numerals.

5 The so-called 'set-A markers' are person markers co-indexed with possessors, agents of transitive verbs, and subjects of intransitive verbs in the imperfective aspect.

Only certain quantifiers require a classifier in Yucatec Maya: numerals of Mayan origin and the interrogative quantifier *jay* 'how many'. The property that distinguishes these elements from other quantifiers is the expression of Cardinality. Non-cardinal quantifiers do not license the classifier construction; see *ya'ab* 'many/much' in (11a). Note that the difference between *ya'ab* 'many/much' and numerals cannot be accounted for in terms of a difference between numerals selecting atomic and numerals selecting non-atomic nominals: the plural subject agreement at the verb (*-o'ob* 'B.3.PL') implies that the nominal *máak* 'man' is individuated without a classifier and a plural suffix. Furthermore, the classifier construction is restricted to numerals of Mayan origin. In (11b), the numeral of Mayan origin *óox* 'three' is used with a classifier while the numerals of Spanish origin *sieete'* 'seven'/*kwaatro'* 'four' are not used with classifiers (see Lehmann 2010: 4–5 and the same phenomenon in Ch'ol in Bale et al. 2019: 10, Little et al. 2022: 12 and Q'anjob'al in Mateo Pedro 2022: 3). These facts indicate that classifiers are not necessary to individuate the nominals, which should also apply to constructions with Spanish numerals (Lehmann 2010: 3). The use of bare nouns instead of plurals (which is possible in Yucatec Maya but not in Spanish) indicates that these examples are not instances of code-switching (Uth & Gutiérrez Bravo 2018, Gutiérrez Bravo & Uth 2020: 84).

(11) (a) quantifier *ya'ab* 'many/much'

ya'ab	*máak*	*k-u*	*bin-o'ob*	*te'el-o'.*
much/many	person	IPFV-A.3	go-3.PL	there-DIST

'Many persons go there.' (Monforte et al. 2010: 225)

(SPEAKER: male, 41Y; bilingual; Timul; rec. 2006-2008)

(b) numerals of Spanish vs. numerals of Mayan origin

Ti'	*sieete'*	*paal-al*	*t-in*	*tséen-t-aj,*	*kwaatro'*
LOC	seven	child-COLL	PFV-A.1	feed-TRR-CMPL[B.3]	four

xíib,	*óox-túul*	*ko'olel*	(...)	*t-in*	*tséen-t-aj.*
man,	three-CL.ANIM	woman		PFV-A.1	feed-TRR-CMPL[B.3]

'From the seven children that I fed, I fed four men and three women.' (Monforte et al. 2010: 33)

(SPEAKER: female, 54Y; monolingual; Timul; rec. 2006-2008)

Numerals of Mayan origin must appear with a classifier even in the absence of nouns, e.g., in counting as in (12a) or in numeric calculations as in (12b) (see also Ch'ol in Little et al. 2022: 25).

(12) (a) *jum-*(p'éel), ka'a-*(p'éel), óox-*(p'éel), kam-*(p'éel)*
one-CL.UNIT two-CL.UNIT three-CL.UNIT four-CL.UNIT
'one, two, three, four'

(b) *jum-*(p'éel) y-éetel u láak' ka'a-*(p'éel)-e' óox-*(p'éel).*
one-CL.UNIT Ø-with A.3 other two-CL.UNIT-TOP three-CL.UNIT
'One and two is three.'

For the expression of measure with quantifiers that do not license the classifier construction, speakers use complex noun phrases; see Section 3.2 (Lucy 1992: 51, Briceño Chel 1992: 79, Lehmann 1998: 63, Gutiérrez Bravo & Uth 2020: 89). These are pseudo-partitive constructions headed by nouns expressing measure: the person agreement clitic (set-A marker) is co-indexed with the noun and the noun is relationalized; cf. (10). Beyond measure expressions, this construction occasionally appears with sortal classifiers, as in (13b). Presumably, the use of classifiers in these examples is a reflex of bias from the constructions with native numerals since the classifier does not have a contribution to the meaning and can be omitted as in (11b).

(13) (a) *t-in man-aj seeis u wóol-ol sakam*
PFV-A.1 buy-CMPL[B.3] six A.3 ball-REL dough
'I bought six balls of dough.'

(b) *oochoj u túul-ul senyoora-s ts'-u k'uch-ul-o'ob-i'.*
eight A.3 CL.ANIM-REL woman-PL TERM-A.3 arrive-INCMPL-PL-LOC
'Eight women have arrived.' (Monforte et al. 2010: 273)
(SPEAKER: male, 82Y; monolingual; Sabacche; rec. 2006-2008)

If classifiers were for nouns, they would be expected to appear independently of numerals (see Section 3.1). Sortal classifiers are not used independently of numerals in Yucatec Maya, e.g., with demonstratives, as illustrated in (14a), which differs from languages in which classifiers are required by the nouns such as Mandarin, Cantonese (Bale & Coon 2014: 698) or Shan (Little et al. 2022: 24–25).[6] Moreover, in case of embedded Spanish phrases as in

6 Instances of classifiers without numerals are reported for spontaneous data, e.g., *p'éel túunich* (CL.UNIT stone) 'one stone' (Pfeiler 2009: 94, Briceño Chel 1994: 148). These examples are always interpreted as involving the omission of the numeral *jun* 'one' and not as underspecified for number.

(14b), the fact that the embedded phrase is interpreted as atomic in Spanish does not affect the need for a numeral classifier.

(14) (a) *Túun táal le (*túul) máak-o'.*
PROG.3SG come DEF CL.ANIM person-DIST
'This person is coming.'

(b) *Yaan jum-p'éel tren de raspa-i'.*
EXIST one-CL.UNIT train of scrape-LOC
'There is a scrape train there.' (Monforte et al. 2010: 111)
(SPEAKER: male, 67Y; bilingual; Canicab; rec. 2006-2008)

These facts indicate that classifiers are required by the numerals in Yucatec Maya, according to the criteria in Section 3.1. Classifiers appear only with quantifiers that are used in expressions of Cardinality (numerals of Mayan origin, interrogative quantifier). Classifiers are obligatory with these elements – even in the absence of nouns as in (12). Numerals of Spanish origin differ in that they do not need a classifier to be used as cardinal numerals, which equally applies to the use of these lexical elements in Spanish and in Yucatec Maya.

3.4 Numerals roots

The only environment in which numerals of Mayan origin do not require a classifier is their use in compounding (Lucy 1992: 50). The formation of compounds with numerals is productive (numerous elements are listed in the available dictionaries; Barrera Vásquez 1980, Bastarrachea Manzano & Canto Rosado 2003, Bricker et al. 1998) and the meaning of the numerals in these compounds is transparent. Two classes of compounds must be distinguished: if the numeral is not interpreted as cardinal, it is not used with a classifier. In these compounds, the numeral is interpreted as a multiplier, indicating the number of iterations of an event such as 'go' in (15a) or of a property such as 'grandmother' in (15b): a *ka'a-chiich* (two-grandmother) is not 'two grandmothers' but rather a 'double grandmother'.[7] Crucially, when the numerals have a cardinal interpretation (counting the number of entities in a set) in compounding, then they are used with a classifier as in (15c).[8]

7 Note also that inherently cardinal numerals (in languages with such numerals) are not directly used in this type of compounds; e.g., Spanish *dos* 'two' but *bis-abuela* (double-grandmother) 'great grandmother'. This shows that a numeral that is valued as 'cardinal', such as *dos* in Spanish, cannot be used as a multiplier.

8 Our argument does not bear on the precise morphological status of the elements in (15c). They are likely listed as lexical items in dictionaries because their meaning is not entirely com-

(15) (a) *ka'a-bin*
two-go
'go for a second time' (Lucy 1992: 50)

(b) *ka'a-chiich,* *óox-chiich*
two-grandmother three-grandmother
'great-grandmother, great-great-grandmother'
(Bastarrachea Manzano & Canto Rosado 2003: 93, 162)

(c) *ka'a-p'éel* *ich,* *ka'a-p'éel* *óol*
two-CL.UNIT face two-CL.UNIT soul
'hypocrite, undecided' (Bastarrachea Manzano & Canto Rosado 2003: 94)

The facts in this section corroborate the view that Yucatec Mayan numeral roots are not inherently cardinal. Inside compounds, we find numerals with classifiers only if the numerals need to be interpreted as cardinal. In conjunction with the findings of Section 3.3, we conclude that the measure function of numeral roots is not valued as 'cardinal'. Thus, classifiers must be used whenever these roots form cardinal numerals – either as modifiers of nouns as in (16a) or as number words (without a nominal) as in (16b).

(16) Numerals and classifiers in Yucatec Maya

(a) [$\text{Num}_{\mu\text{:unvalued}}$ $\text{Clf}]_{\mu\text{:CARDINAL}}$ N
jun *p'éel* *tuunich*
one CL.UNIT stone

(b) [$\text{Num}_{\mu\text{:unvalued}}$ $\text{Clf}]_{\mu\text{:CARDINAL}}$
jun *p'éel*
one CL.UNIT

Some Mayan languages have an affix for the formation of numerals that does not have the function of a classifier; see Kaqchikel García Matzar & Rodríguez Guaján (1997: 137, 155) and Tzeltal (Berlin 1968: 23, Polian 2013: 595). In Tzeltal, the suffix *-eb* 'NUM' is attached to numeral roots rendering cardinal numerals that can be either used as number-denoting words (e.g., in counting) or as modifiers of nouns (without a classifier). Alternatively, the nominal roots combine with specific classifiers that identify certain classes. This gives rise to two alternative expressions: four pigs' can be expressed through *chan-eb chitam* (four-NUM pig) or through *chan-kojt' chitam*

positional, as it arises from synecdoche. The crucial issue is that in expressions involving a cardinal interpretation, numbers must appear with classifiers.

(four-CL.ANIMAL pig) (Polian 2013: 595). This suffix originates in a plural marker in Tzeltal (Polian 2013: 595) and Kaqchikel (García Matzar & Rodríguez Guaján 1997: 137), while the corresponding expression in Yucatec Maya originates in a classifier. Hence, different Mayan languages use elements of different origin in order to form cardinal numerals out of numeral roots that are not intrinsically cardinal.

Recall that according to the 'classifiers for numerals' account in (9), the presence of classifiers does not imply the absence of a count-mass distinction (see Section 3.1). Yucatec Maya has nouns that can be used with atomic denotation without a classifier; see individuation of *ya'ab máak* 'many persons' with a vague quantifier in (11). Nouns of substances are typically used without atomic denotation to the effect that the same quantifier would not quantify the number (of individuals), but the volume (of a substance): *ya'abach ixi'im* 'much corn' (Monforte et al. 2010: 25). The same contrast applies to the use of plural: nouns with atomic denotation can express sums of atomic entities if they are pluralized, e.g., *máak-o'ob* (person-PL) 'persons'; nouns denoting substances can be pluralized if the contextual cues allow for coercion to atomic entities, e.g., *lu'um-o'ob* (soil-PL) 'countries' (Yu 2024: 106–108).

3.5 Classifiers and measure expressions

Expressions of measure are used to denote quantitative units: they may apply to masses, e.g., 'one gourd of water' (the unit measures a certain quantity), to pluralities, e.g., 'one herd of cows' (the unit is a certain sum of atoms), or to atomic entities that can be conceptualized as homogeneous, e.g., 'one piece of watermelon' (the unit measures a part of the entity). The shared property of these types of nominal denotation is a notion of homogeneity, which means that the quantified entities are (a) cumulative (any sum of x is x) and (b) divisive (any part of x is x) (Deal 2017: 129).

There are two types of expressions of measure in Yucatec Maya. Some expressions of measure are ordinary nouns, such as the Spanish borrowing *kiiloj* 'kilo' in (17). As all further nouns, these terms are used in numeral constructions with the general classifier (Briceño Chel 1993: 159, Briceño Chel 1994: 147); the same phenomenon is reported for Ch'ol (Bale et al. 2019: 25) and Chontal (Knowles-Berry 1987: 336). In the following, we use the term 'measure nouns' for nouns expressing units of measurement when they are used in numeric expressions.

(17)	*p'is*	*ten*	*óox-*(p'éel)*	*kiiloj*	*bu'ul!*
	weigh[IMP]	1SG	three-CL.UNIT	kilo	bean

'Weigh three kilos beans for me!' (Briceño Chel 1994: 147)

Mensural classifiers have a different behaviour, namely they can be directly juxtaposed to the numerals, as in (18).

(18) (a) *jun-xúuxak* *je'-o'ob*
one-CL.BASKET egg-PL
'one basket of eggs'

(b) *jun-xéet'* *tuunich*
one-CL.PIECE stone
'one piece of stone'

Mensural classifiers have a dual nature. Similarly with measure nouns, they express units of measuring masses, pluralities or homogeneous atomic entities that are counted in numeric constructions. At the same time, they can attach to numeral roots without the mediation of the general classifier (in contrast to measure nouns).[9]

Finally, mensural classifiers differ from sortal classifiers. By definition, sortal classifiers identify the class of the nominal predicate while mensural classifiers denote a unit of measurement (Croft 1994). Hence, substituting a sortal classifier by a general classifier is not truth-conditionally relevant, e.g., *jum-p'éel che'* 'one-CL.UNIT tree' has the same extension with *jun-kúul che'* 'one-CL.PLANT tree'. This does not apply to mensural classifiers since the unit of measurement has a contribution to the meaning. The expression *jum-p'éel ja'* (one-CL.UNIT water) 'one water' does not have the same extension with *jun-luuch ja'* (one-CL.GOURD water), since the option with the general classifier can refer to any portion of water that can be conceptualized as atomic in a given context, as in (19).

(19) Context: 'As for the annatto, you shake it well...'

k-a *beet-ik* *jum-p'éel* *nuxi'* *ja'* *nojoch-e'*
IPFV-A.2 make-INCMPL[B.3] one-CL.UNIT huge water big-CONT
'you make one huge water' (Skopeteas et al. 2020: YUC-TXT-CO-21)

(SPEAKER: F; 22Y; bilingual; F. Carrillo Puerto; rec. 2015)

3.6 *Reassessing the sources of microvariation*

The previous sections discussed four types of terms used in numeral expressions; see (20). Sortal classifiers denote the 'sort' of the nominal predicate,

9 In this sense, mensural classifiers are a type of head defined syntactically; these elements are not necessarily measures when used outside numeral expressions.

while the general classifier applies to all 'sorts' of nominal predicates. General, sortal, and mensural classifiers can be directly juxtaposed to the numeral root whenever it is used as a cardinal number. Measures can also be expressed by nouns that are attached to the numeral through the mediation of a classifier.

(20)

expression	CARDINALITY	SORT	MEASURE
general classifier (*p'éel* 'CL.UNIT')	√	–	–
Sortal Classifier (*kúul* 'CL.PLANT')	√	√	–
Mensural Classifier (*wóol* 'CL.BALL')	√	–	√
Measure Noun (*kiiloj* 'kilo')	–	–	√

In light of (20), the change from a sortal to a general classifier is an instance of desemanticization: the sortal classifier loses the feature of denoting a specific 'sort'. The mediation of the general classifier in numeric expressions of measure involves two processes: (a) mensural classifiers lose their dual nature, attaching to numerals with the mediation of the general classifier (just as measure nouns); (b) the use of the classifier *p'éel* in these expressions leads to the obligatorification of this classifier in all uses of numeric roots as cardinal numerals.

The evolution of the general classifier is complemented by developments in the specific (sortal/mensural) classifiers. The paradigm of sortal classifiers shrinks over time (Thompson 1970: 319, Briceño Chel 1992: 71–75, Bricker 2019: 244), which is generally part of the diachronic process leading to the loss of classifier systems (Grinevald 2000: 85–86). The use of multiple-classifier constructions (containing the general classifier) implies that mensural classifiers lose their dual nature and turn to measure nouns, which is a process attested in the history of Yucatec Maya (see Section 2).

4 Method

Our aim is to assess the relation between two current processes in the numeral-classifier constructions in Yucatec Maya: (a) the substitution of sortal classifiers by the general classifier and (b) the use of mensural classifiers with the general classifier. The resource that we are using is presented in Section 4.1, the data of the present study in Section 4.2. Section 4.3 introduces the factorial design, Section 4.6 the statistical procedures. Our predictions for the outcome are presented in Section 4.4 and the threads to the validity of our inferences in Section 4.5.

4.1 Resource and methodological considerations

The current study uses data from the *Atlas of Yucatec Maya* (referred to as AYM henceforth), a resource containing data from a sample of locations across the Yucatecan peninsula (Blaha Pfeiler 2021). This dataset was gathered between 2000 and 2007 at the *Universidad Autónoma de Yucatán* in Mérida, Mexico. The data collection took place in 86 locations chosen to represent regions that were reported to display dialectal differences in earlier studies (Briceño Chel 2002, Bastarrachea Manzano & Canto Rosado 2003, Pfeiler 2014: 208—209): Northeast, Northwest, Southern Yucatán, Center of Quintana Roo, the areas Los Chenes and Camíno Real in the state of Campeche as well as an enclave of Yucatec Maya in Belize (see maps in Section 5). The data-gathering team sought for maximally competent speakers in the locations at issue and created a corpus of 176 speakers (117 female; year of birth ranging between 1906 and 1989, mean: 1953, median: 1953). Interviews were recorded and transcribed by a team of native speakers of Yucatec Maya; transcriptions, sound files and additional information are accessible online (Blaha Pfeiler et al. 2022).

The data collection encompasses prompts designed to assess variation in different linguistic layers (phonology, morphology, syntax, lexicon) (total $n = 665$). The empirical design has two basic properties that were uniformly used across trials: it is elicited through *translation* and relies on *minimal prompts* (without contextual information, also using subclausal constituents in various sections).

Translation tasks are widely used in dialectological data collections (Bucheli & Glaser 2002, Cornips 2002, Brandner 2012). A translation task constitutes a complex endeavor (Bohnemeyer 2015: 20, 25), involving:

- the comprehension of the prompt in the source language, and
- the choice among alternative expressions in the target language.

During the comprehension process, the speaker identifies the extension of the prompt. This is a crucial difference between expressions elicited with non-verbal stimuli (e.g., through picture-based descriptions) and expressions elicited through translation. In the former case, the targeted content is a depicted situation, in the latter it is the extension of a linguistic expression (in the source language). This means that variation may emerge in translation tasks, especially whenever the prompt in the source language is underspecified or ambiguous (Matthewson 2004: 391–393). Additionally, the data of the present resource rely on minimal prompts that are presented 'out-of-the blue'. Thus, potential ambiguities of the source expression are not restricted by the sentential environment or the context, which may lead to an additional increase of variability, since speakers may imagine different contexts during the comprehension of the prompt (see Tonhauser & Matthewson 2016: 15 and Bochnak & Matthewson 2020: 263).

During the production process of the translation task, the speaker has to identify an expression in the target language that has a maximal overlap with the extension of the prompt. The choice of response can be modulated by at least two further factors: expressions of the target language that partially fit to the prompt may differ in salience, which may bias towards a response that is readily accessible in memory. These biases, if not random, may provide valuable insights: dialects not only vary in the presence of certain variants but also in their preference for certain variants – which may otherwise be present across dialects. Furthermore, the outcome of the translation can be influenced by interferences from the language of the prompt. Properties of the prompt might prime for an expression in the target language that would not be favored outside the translation context (Bucheli & Glaser 2002: 43, Brandner 2012: 123, Cornips 2002: 91, Bohnemeyer 2015: 20, 25). The relevance of these considerations for the validity of our inferences from the data is discussed in Section 4.5.

4.2 *Data*

In the present study, we examine 23 prompts that are expected to elicit expressions with numeral classifiers; see summary in Table (1) and full listing in Appendix I. The AYM data contains 4 prompts for sortal classifiers that are informative for the choice between the general and a specific classifier.[10] Furthermore, the data provides three groups of prompts (n=19) for measure expressions, grouped under *Portions* of substances, *Sums* of pluralities, and *Parts* of homogeneous entities. These categories were included in AYM in order to provide an observational basis with diverse types of classifiers (based on Miram 1983) and are not supposed to be exhaustive.

Due to missing values, the total dataset comprises 3446 responses by 173 speakers, which corresponds to 86.6% of the 3979 possible permutations of 173 speakers × 23 prompts. Informative tokens for the use of general classifier are responses such as (21) that contain a numeral, one or more classifiers and a noun (see further examples in Section 5).

10 A fifth prompt of this group, namely 'one stone' (Q164) is not used in the present study because the responses do not involve any variation (the classifier *p'éel* was selected by all speakers).

group	examples	*n*
Sorts:	'one man', 'one candle', 'one stone', 'one plant of huaya'	4
Portions:	'one ball of dough', 'one fistful of cement', 'one pile of excrement', 'two shots of water', 'one drop of water'	5
Sums:	'one row of stones', 'one pile of paper', 'one load of firewood', 'one bunch of herbs', 'one bundle of clothes', 'two weights of beans', 'one mooring of beans'	7
Parts:	'one half of watermelon', 'one piece of stone', 'one cut piece of wood', 'one broken piece of wood', 'one piece of clothing', 'one torn piece of cloth', 'two slices of cheese'	7

Table 1 Groups of prompts in AYM

(21) PROMPT: *una bola de masa* 'one/a ball of dough' (Q167)

jum-p'éel *wóol* *sakan*
one-CL.UNIT CL.ROUND dough
'one ball of dough' (Q167-S113)[11]

(SPEAKER: male, 61Y; bilingual; Texán de Palomeque; rec. 2007)

Responses that are not informative for our research question were classified as 'non-valid'. The following conditions were used for selecting the valid data: (a) CONTENT AT ISSUE: the response lexicalizes the targeted content, which excludes various types of deviation from the meaning of the prompt (*n*= 359), as for instance the use of non-numeric expressions, e.g., prompt 'one bunch of herbs', response *ya'abkach xíiw* 'many/much herb' (Q181-S044) or cases in which the targeted measure of the prompt is not expressed in the response (*n*= 29), e.g., prompt 'two portions of beans', response *ka'a-p'éel bu'ul* (two-CL.UNIT bean) 'two beans' (Q184-S144); (b) CONSTRUCTION AT ISSUE: the response realizes the content with the targeted construction, excluding responses that do not involve a native classifier construction and a

11 AYM codes: question identifier (Q167) – speaker identifier (S113).

noun (n= 435), such as expressions of measure that are not classifiers (n= 49), e.g., prompt 'one row of stones', response *jun-p'éel tsool-bil tuunich* (one-CL.UNIT line.up-ADJR stone) 'one unit of rowed stones' (Q175-S169), responses without an overtly realized head noun (n= 225), e.g., prompt 'two slices of cheese', response *ka'a p'aay* (two CL.CRUMB) 'two crumbs' (Q183-S169), and responses with syntactic priming from Spanish (n= 9), e.g., prompt 'one row of stones', response: *jum-p'éel chan piilaj de tuunich* (one-CL.UNIT small line of stone) 'just one row of stones', which includes the Spanish preposition *de* 'of' (Q169-S170). Nouns of Spanish origin are occasionally used in measure expressions (n = 152). As already discussed in in Section 3.3, Spanish nouns cannot be directly attached to numeral roots, but are only used with the mediation of the general classifier. Since this expectation is confirmed by the AYM data, we restrict our analysis to the elements of Mayan origin.

Four speakers (totaling 6 responses) were excluded because they did not produce any valid responses either in the prompts of sortal or in the prompts of mensural classifiers. Subsequent analyses examine the valid subset of the data of 169 speakers (from 85 locations), which comprises 2645 responses (76.8% of 3446 available responses). When speakers provided multiple responses to the same prompt, often due to self-correction, we evaluated only the final valid response. The number of valid responses per prompt ranged from 64 to 166 (out of 169 speakers), with an average of 115 (median: 127) valid responses; see Figure 1 (LEFT PANEL). On average, speakers contributed (mean/median) 16 valid responses out of 23 prompts, with valid response counts ranging from 4 to 23; see Figure 1 (RIGHT PANEL).

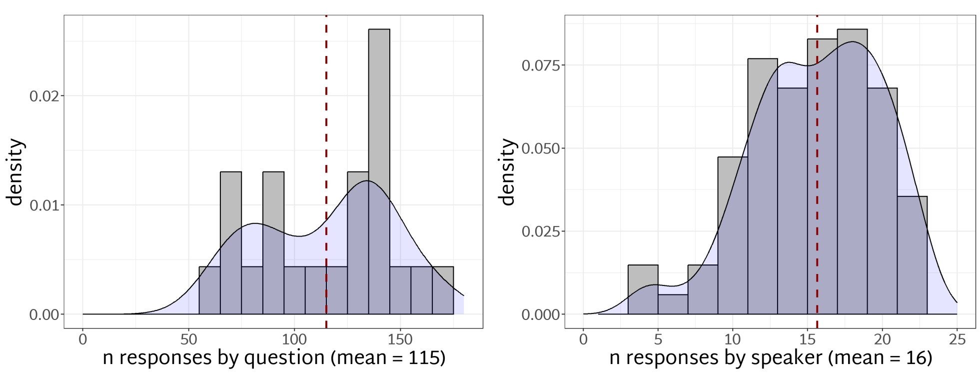

Figure 1 Histograms of the valid data per question (LEFT PANEL) and per speaker (RIGHT PANEL); blue filled area indicates density, red dashed line shows the mean.

4.3 Factors

The phenomena introduced in Section 1 are treated in two separate analyses, both of which use the factor CONSTRUCTION as the dependent variable. For sortal classifiers, we examine the substitution of a (specific) sortal classifier by a general classifier. For mensural classifiers, we examine the use of a general classifier in combination with a mensural classifier. Hence, the factor CONSTRUCTION has two levels in either analysis:

- 'Specific classifier' construction: 'numeral + specific sortal classifier' as in (3a-c) (in the sortal-classifiers model in Section 5.1); 'numeral + mensural classifier' as in (5) (in the mensural-classifiers model in Section 5.2);
- 'General classifier' construction: 'numeral + general classifier' as in (4) (in the sortal-classifiers model in Section 5.1); 'numeral + general classifier + mensural classifier' (multiple-classifier construction), as in (6) (in the mensural-classifiers model in Section 5.2).

The following factor is used to assess preferences that are explained by the speakers' biases from Spanish:

- SPANISH BIAS: Spanish/Mayan ratio in the total responses of each speaker in the AYM questionnaire; range between 0 (i.e., no Spanish words) and 0.2 (i.e., 2 Spanish words : 10 Yucatec Mayan words); based on the annotation of (total) 31,563 responses in 237 prompts of the entire corpus, that contain 30,273 Mayan words (96%) and 1,268 Spanish words (4%).

The following factors are used to assess the dispersion of these constructions in the population:

- TIME: apparent-time scale of the speakers' year of birth: range between 1906 and 1989 (see Bailey, Wikle, Tillery & Sand 1991, Bailey 2004 on apparent-time scales).
- SPACE: contains the geographical coordinates (latitude, longitude) of each speaker's location, modelled as a non-linear factor (see Section 4.6 on statistical modelling).
- POPULATION SIZE: logarithmized population size of the locations, based on the closest census to the time of the data collection (INEGI 2011); range between 3.7 (44 speakers, Bombahaltún) and 12.3

(220,389 speakers, City of Campeche); logarithmization is due in order to avoid effects of the power-law distribution of population sizes (Wieling, Nerbonne & Baayen 2011, Wieling, Montemagni, Nerbonne & Baayen 2014).

In the analysis of mensural classifiers, the following factors were added to detect dependencies between linguistic categories:

- Unit/Sort: likelihood of the general classifier [0-1] to substitute a sortal classifier per speaker – to assess whether the use of general classifier in sortal-classifier constructions predicts the use use of general classifier with mensural classifiers.
- Measure Type: categorical factor with three levels (*Portions*, *Sums*, *Parts*), contrasts modelled as successive differences – to assess variation due to the different types of measure expressions.

4.4 Predictions

The variation between the classifier constructions may result from two diachronic processes, which are not mutually excluded:

(22) Hypotheses about language change

(a) Change in the general classifier
The classifier *p'éel* evolves into a general marker of *Cardinality*.

(b) Change in the mensural classifiers
The mensural classifiers lose their dual nature and are only used as measure nouns.

If the change lies in the general classifier as in (22a), then both phenomena (the generalization of the classifier with sortal classifiers and the use of the general classifier with mensural classifiers) are traced back to a common process of change. This implies that the likelihood of substituting a sortal classifier with the general classifier should be a significant predictor for the likelihood of using the general classifier with mensural classifiers. The change in mensural classifiers in (22b) does not make any predictions about the correlation of these phenomena.

Furthermore, the hypotheses in (22) make different predictions about the impact of Measure Type. If the change relates to the general classifier as in (22a), it is expected to uniformly apply to all types of measure. If the change lies in the properties of mensural classifiers as in (22b), the effect of potential changes may vary between measure types.

The hypothesis in (22a) predicts that the same parts of the population that prefer the general classifier in sortal-classifier constructions will also prefer the general classifier in mensural-classifier constructions. This prediction equally applies to all demographic factors: areal groups (effects of Space), age groups (effects of Time), and groups determined by the rural-urban dimension (effects of Population Size).

The predictions for the distributions in Space follow the reasoning of microvariation studies seeking for implicational relationships between grammatical phenomena across dialectal varieties (see, for instance, Poletto 2000: 134, Brandner 2020: 26). Dependencies between grammatical properties are expected to show up in a subset relation in geographical space, as schematically sketched with two properties α and β in Figure 2(a). In the subset scenario, property β implies property α ($\beta \rightarrow \alpha$). If the properties vary independent of each other, then the sets of locations in which the properties α and β may be either disjoint, as in Figure 2(b), or having a partial overlap, as in Figure 2(c). In either scenario, the properties α and β do not depend on each other; see, e.g., negative doubling and negative spread in Moser (2019). Partial overlaps may be informative if the intersection is large enough to justify the relationship. A residual lack of overlap can be accounted for if certain properties spread outside the expected area through language contact; see, e.g., Fleischer (2000: 132) about the partial overlap of dialectal areas of German regarding the stranding and the long doubling construction of prepositions.

Applying this reasoning to our data: the hypothesis (22a) predicts the pattern in Figure 2(a), whereby α stands for the use of the general classifier in sortal constructions and β for the use of the general classifier in mensural constructions. In terms of the hypothesis in (22b), these phenomena are independent from each other, which does not make any predictions about their distribution in space.

Finally, the choice of Construction in the responses may involve various effects such as random spontaneous decisions, task-specific effects, and biases of Spanish. Task-specific effects (e.g., influence by the interview context, the translation task) as well as accidental spontaneous choices of the individuals are part of the residual variation, namely the variation that is not captured by the examined factors. Influences of Spanish are expected to be captured by the factor Spanish Bias which reflects the preference of speakers to use elements from Spanish in this type of task.

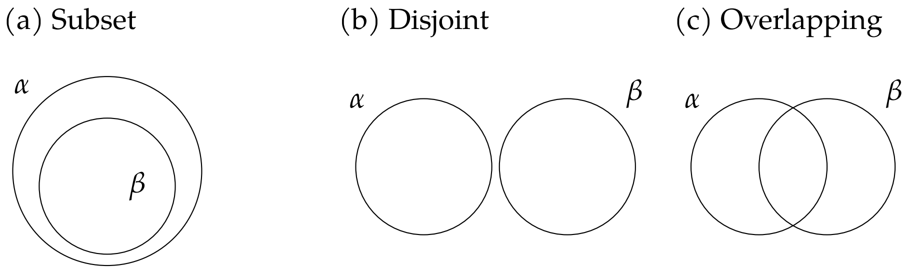

Figure 2 Dispersion of two grammatical properties in space

4.5 Validity issues

Lastly, we need to estimate the validity of our inferences in the light of the properties of this data: (a) expressions were obtained through translation from Spanish; (b) elicitation relied on minimal prompts, lacking a richer context (see Section 4.1). Our study examines whether the conditional probability of the general classifier in two constructions (sortal and mensural classifiers) varies between varieties of Yucatec Maya (captured by sociodemographic variables) (see details in Section 4.3).

The *internal validity* of our inferences (i.e., inferences from the data to the statements) can be threatened from potential confounds between artefacts of the elicitation method and the independent variables. Since the elicitation method was kept consistent across speakers, no confound applies to this type of inferences. A potential thread is the confounding of population groups with the degree of Spanish bias. This thread is treated in the factorial design, since it includes the relevant predictor that is expected to capture the speakers' inclination to use Spanish expressions in this type of task (Spanish bias). The lack of context and the uncertainty due to potential ambiguities of Spanish prompts predict an increase of variability in the data; this implies an increase of the residual variation, i.e., the variation that is not explained by the factors included in the statistical model (see Section 4.4).

The *external validity* of our inferences (i.e., inferences from the sample to the population) is more difficult to assess. It is obvious that the controlled interview setting in general, the translation task itself as well as the absence of context threaten the generalizability of the findings to real-life situations. The use of translations from Spanish may lead to an increase of using the general classifier, since Spanish numeral expressions do not have (sortal) classifiers.

The absence of context in the prompts may involve a bias towards underspecified expressions, i.e., speakers may opt for a general classifier whenever they have the choice between the general classifier and a specific one. Although the general intuition about the Spanish bias is reasonable, the translation task itself biases towards expressions of the target language. For instance, the proportion of Spanish words in the AYM data (4% Spanish words in 31,541 annotated responses) is lower than the proportion of Spanish words in the spontaneous conversations in Yucatec Maya (9% Spanish words out of 13,345 words; reported by Pool Balam & Le Guen 2015: 372). With this background, the *overall frequency* of the general classifier in this data is not generalizable for all types of language use (see comparison between translation tasks and spontaneous data in Cornips 2002). However, our study does not draw inferences from the overall frequency of the general classifier, but from the influence of Space, Time, Population Size on the its use. Crucially, the effects of interest are not confounded with the threads at issue.

4.6 Statistical Analysis

All linear factors (Time, Population Size, Spanish Bias) were rescaled into an interval between .001 and .999, in order that the magnitudes of the estimates are comparable (Wieling et al. 2011: 5). The occurrence of 'general classifier' constructions was fitted by the models in (23) in two separate analyses that assess the co-efficients β of the independent variables presented in Section 4.3. In order to reduce complexity, we only test the interaction effect that is motivated by the hypothesis in (22b), namely the interaction between Space and Measure Type in the analysis of mensural classifiers. These models contain the non-parametric coefficients of Space, which cannot be assessed by a linear model.

(23) (a) Model I: sortal classifiers

Prob(Construction= 'general classifier') =

$\beta_{\text{Intercept}} + \beta_{\text{Spanish Bias}} + \beta_{\text{Time}} + \beta_{\text{Pop. Size}}$ parametric

$+ \beta_{\text{Space}}$ non-parametric

(b) Model II: mensural classifiers

Prob(Construction= 'general classifier') =

$\beta_{\text{Intercept}} + \beta_{\text{Spanish Bias}} + \beta_{\text{Time}} + \beta_{\text{Pop. Size}}$

$+ \beta_{\text{Unit/Sort}} + {}_{\text{Measure Type}}$ parametric

$+ \beta_{\text{Space:Measure Type}}$ non-parametric

In order to estimate the non-parametric factor (i.e., the effect of Space), we used *Generalized Additive Models* (GAM), which assess the factorial effects by fitting non-parametric smoothers (Wood 2017); for linguistic applications, see Baayen & Linke 2020). GAMs have been applied to assess dialectal variation in space by various studies (Wieling et al. 2011, 2014, Wieling, Valls, Baayen & Nerbonne 2018, Wolk & Szmrecsanyi 2018, Blaha Pfeiler & Skopeteas 2022).

The significance of the effects at issue was assessed with model comparison, based on a backwards-elimination procedure: starting with the maximal models in (23), we stepwise compared each model with a model in which a factor of interest was removed. The reduced model was selected unless the larger model reached a smaller AIC (= Akaike Information Criterion) value, significantly improving the model deviance (i.e., the difference between model deviances corresponded to a p-value below .05 in the chi-square distribution). Calculations were made in R (R Core Team 2023), using library *mgcv* (Wood 2010) for generalized additive models and library *itsadug* for model comparisons (van Rij, Wieling, Baayen & van Rijn 2022).

5 Results

5.1 *Sortal classifiers*

The subset of valid data with sortal classifiers comprises 576 responses (elicited with four different prompts; see Appendix I/A). In most cases (53%), the numeral expression is formed with the specific classifiers *túul* 'cl.anim' (for 'one man' in Q111 and 'one pig' in Q163), *kúul* 'cl.plant' (for 'one plant' in Q165), and *ts'íit* 'cl.long' (for 'one candle' in Q166), as in (24a). Many responses (45%) use the general classifier *p'éel* 'cl.unit', as in (24b). Finally, a few responses (1.4%) involve the use of general classifier with a sortal classifier as in (24c), which is a construction that rarely appears in our data (and never with the classifier *túul* 'cl.anim').

(24) (a) Sortal-classifier construction
(n = 307; 53.3% out of total 576)

jun-ts'íit *kib*
one-cl.long candle
'one/a candle' (Q166-S160)

(Speaker: male, 82Y; bilingual; Kancabdzonot; rec. 2003)

(b) General-classifier construction
(n = 261; 45.3% out of total 576)

jum-p'éel *kib*
one-cl.unit candle
'one/a candle' (Q166-S126)

(Speaker: male, 81Y; bilingual; Pixoy; rec. 2004)

(c) Multiple-classifier construction
(n = 8; 1.4% out of total 576)

jum-p'éel *kúul* *wayáam*
one-CL.UNIT CL.PLANT huaya
'one/a huaya plant' (Q165-S161)

(Speaker: male, 53Y; bilingual; Nah Balam; rec. 2006)

The following analysis only examines the choice between the specific sortal classifier as in (24a) and the general classifier as in (24b) (n=568). The proportions of expressions with the general classifier are plotted by location in Figure 3/Left Panel. These proportions range between 0% (no use of the general classifier in the four prompts for sortal classifiers) and 100% (use of the general classifier in all four prompts).

A generalized additive model (GAM) was fitted to the choice of Construction in the prompts for sortal classifiers with the parameters in (23a). The estimates of the model of best fit are listed in Table 2. The parametric coefficients of this model reveal a significant main effect of Spanish Bias, with a positive estimate (= .455), indicating that the probability of using the general classifier increases with speakers who display a larger Spanish bias in their overall responses in the AYM data. Time and Population Size could be removed without significantly reducing the informativity (AIC) of the model.

	parametric coefficients				model comparison	
	β	SE	t	$p(>\|t\|)$	χ^2	$p(>\|\chi^2\|)$
Intercept	.359	.033	11.005	<.001		
Spanish Bias	.455	.119	3.811	<.001	6.939	<.001
Time					1.398	n.s.
Population Size					.019	n.s.
	edf	F	$p(>\|F\|)$			
Space	11.83	3.818	<.001			

Table 2 GAM on the occurrence of general classifier in sortal-classifier constructions (n = 568; Deviance explained = 15.7%)

The factor Space is treated differently in Table 2, since it is a non-parametric factor (see Section 4.6). The coefficients of Space are presented in Figure 3/Right Panel in a physical-map visualization, such that the level of prediction of the general classifier is mapped onto elevation levels. These coefficients reveal that the variation in the data is optimally captured by a model displaying an increase of the use of general classifiers in the West/North. The effective degrees of freedom (edf= 11.87) in Table 2 indicate that the model of the best fit substantially deviates from a model with a linear term for SPACE, which justifies the treatment of this factor as non-linear.[12] The F-statistic shows that the effect of SPACE is significant.

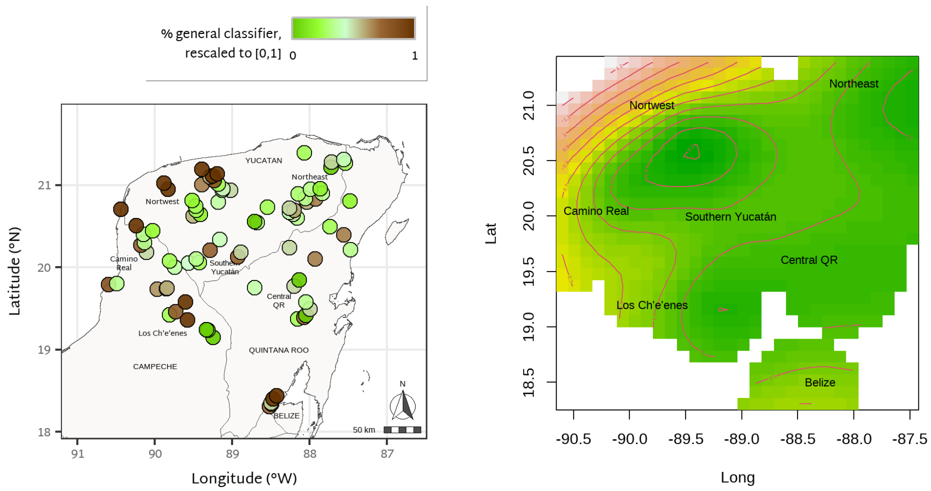

Figure 3 Sortal Classifiers in SPACE: LEFT PANEL: % general classifier out of total valid responses in the prompts for sortal classifiers; RIGHT PANEL: GAM-coefficients of SPACE on the occurrence of the general classifier; level of prediction mapped onto the elevation colors of physical maps: green < brown < white.

5.2 *Mensural classifiers*

The subset of valid data with mensural classifiers comprises 2072 responses (elicited through 19 prompts; see APPENDIX I/B-D). In most valid responses (86.3%), mensural classifiers directly accompany the numeral, as in (25a); in the remaining responses (13.7%), the mensural classifier attaches to the numeral with the mediation of the general classifier, as in (25b) (mensural classifiers may be more than one).

12 The effective degrees of freedom (edf) are an estimate of the non-linearity of the factor: a value of 1 means that the effect is linear, a high value (8 or higher) means that the effect is highly non-linear (Zuur, Ieno, Walker, Saveliev & Smith 2009: 55).

(25) (a) Mensural-classifier construction
(n = 1786; 86.3% out of total 2069)

jun-kúuch *si'*
one-CL.LOAD wood
'one/a load of wood' (Q179-S153)

(SPEAKER: female, 40Y; bilingual; Isla Arena; rec. 2006)

(b) Multiple-classifier construction (i.e., general+specific classifier)
(n = 283; 13.7% out of total 2069)

jum-p'éel *kúuch* *si'*
one-CL.UNIT CL.LOAD wood
'one/a load of wood' (Q179-S113)

(SPEAKER: male, 61Y; bilingual; Texán de Palomeque; rec. 2007)

Expressions of measure in complex noun phrases (n = 22) are excluded from this data, since they are not instances of the classifier construction (see Section 4.2); see (26), cf. (21).

(26) *jun-p'éel* *u* *wóol* *xa'ak'*
one-CL.UNIT A.3 ball dough
'one/a ball of dough' (Q167-S112)

(SPEAKER: female, 52Y; bilingual; Cobá; rec. 2006)

The relative frequency of using the general classifier (in multiple-classifier constructions) varies between MEASURE TYPES; see Table 3. This construction is most frequent with *Sums*, less so with *Portions* and even less so with *Parts*.

MEASURE TYPE	Mensural		Multiple		Total	
	n	%	n	%	n	%
Parts	*819*	96.1	*33*	3.9	*852*	100
Portions	*410*	83.2	*83*	16.8	*493*	100
Sums	*557*	76.9	*167*	23.1	*724*	100

Table 3 Frequency of classifier constructions by MEASURE TYPE

Since there are no independent reasons predicting the effect of MEASURE TYPE, it is possible that the frequencies in Table 3 emerge through purely lexical differences: some lexical elements are more commonly used as classifiers (directly attaching to the numeral), while other lexical elements normally require the general classifier in numeric expressions. A plausible expectation is that frequent mensural classifiers are more likely to attach directly to the numeral. In order to test this hypothesis, we plot the frequencies of individual classifiers along with the proportions of multiple-classifier constructions with the same elements; see Figure 4 (see lists of mensural classifiers and frequencies in APPENDIX II/B-D).

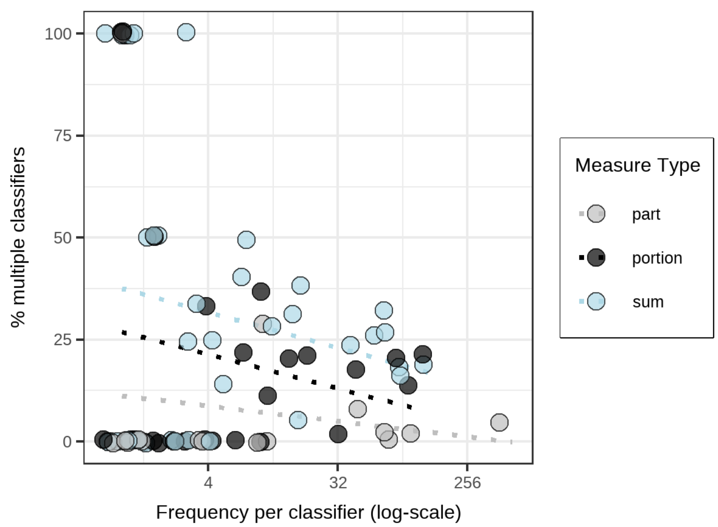

Figure 4 MEASURE TYPES: % multiple classifiers out of total occurrences (y-axis) and n occurrences (x-axis) of each mensural classifier; dotted lines display trendlines per group

Figure 4 reveals that the most frequent classifiers are less likely to appear with the general classifier, which means that they directly attach to the numerals more often than rare classifiers. Hence, the large difference between *Parts* and other measure types may be just due to the fact that most prompts of this measure type were rendered with the mensural classifiers *xéet'* 'CL.PIECE' (n=503 out of 851) and *xóot'* 'CL.CUT_PIECE' (n=106 out of 851), which rarely appear with the general classifier in this data (see Appendix II/C). This dataset cannot be used to test the geographical dispersion of individual lexemes, since many classifiers are too marginally represented to examine their exact distribution. Therefore, the following analysis can only rely on the aggregated results per MEASURE TYPE. However, it is relevant for the interpretation that the likelihood of using the general classifier (in a multiple-classifier construction) in groups of prompts (MEASURE TYPES) may just reflect the variation between the lexical items of the respective group – especially since there are no independent linguistic reasons that would predict the difference between Measure Types.

A GAM was fitted to the choice of Construction in the prompts for mensural classifiers with the parameters in (23b); see coefficients in Table 4. The parametric co-efficients contain a significant main effect of Unit/Sort, indicating that the speaker's preference to substitute sortal classifiers by the general classifier is a significant predictor for the choice of multiple-classifier constructions in expressions of measure. The further factors have similar effects with the analysis of sortal classifiers (cf. Table 2): the significant effect of Spanish Bias indicates that speakers who used more Spanish in their overall responses in AYM are more likely to choose the multiple-classifier construction in measure expressions. Time and Population Size did not reach significance. Furthermore, both successive differences between Measure Types obtained significant effects (confirming the differences in Table 3). The threefold distinction (*Parts* vs. *Portions* vs. *Sums*) was found to have a better fit than models containing a twofold distinction, either between *Parts* and *–Parts* (decrease of the ML score: 5.371, $p < .01$) or between *Sums* and *–Sums* (decrease of the ML score: 30.177, $p < .001$).

	parametric coefficients				model comparison	
	β	SE	t	$p(>\|t\|)$	χ^2	$p(>\|\chi^2\|)$
Intercept	.058	.013	4.386	<.001		
Unit/Sort	.096	.025	3.813	<.001	7.188	<.001
Measure Type					80.015	<.001
Sums – Portions	.061	.018	3.276	<.01		
Portions – Parts	.141	.018	7.808	<.001		
Spanish Bias	.242	.043	5.632	<.001	14.559	<.001
Time					.034	n.s.
Population Size					0	n.s.
	edf	F	$p(>\|F\|)$			
Space:Parts	2.001	5.102	<.001			
Space:Portions	5.849	2.755	<.001			
Space:Sums	11.379	7.071	<.001		12.337	<.001

Table 4 GAM on the occurrence of general classifier in mensural-classifier constructions ($n = 2069$; deviance explained: 16.2%)

The results reveal a significant interaction between Space and Measure Type (see χ^2 value in Table 4). The coefficients of the best-fit model are plotted per Measure Type in Figure 5, while the percentages of general classifier across Measure Types are plotted per location in Figure 5/Top Left Panel. The dispersion of the general classifier of all Measure Types are captured by models with a higher elevation in the East/North, which differs from sortal classifier constructions (cf. Figure 3). Beyond the common trend in the East-West axis, the groups of prompts differ significantly from each other: the probability of using the general classifier with *Parts* (Top Right Panel) can be captured by a nearly linear model, East vs. West (edf value close to 1 in Table 4), while the dispersion of *Sums* and *Portions* (Bottom Panels) are optimally fitted by more complex models containing – among else – an increase of the use of general classifiers in the North.

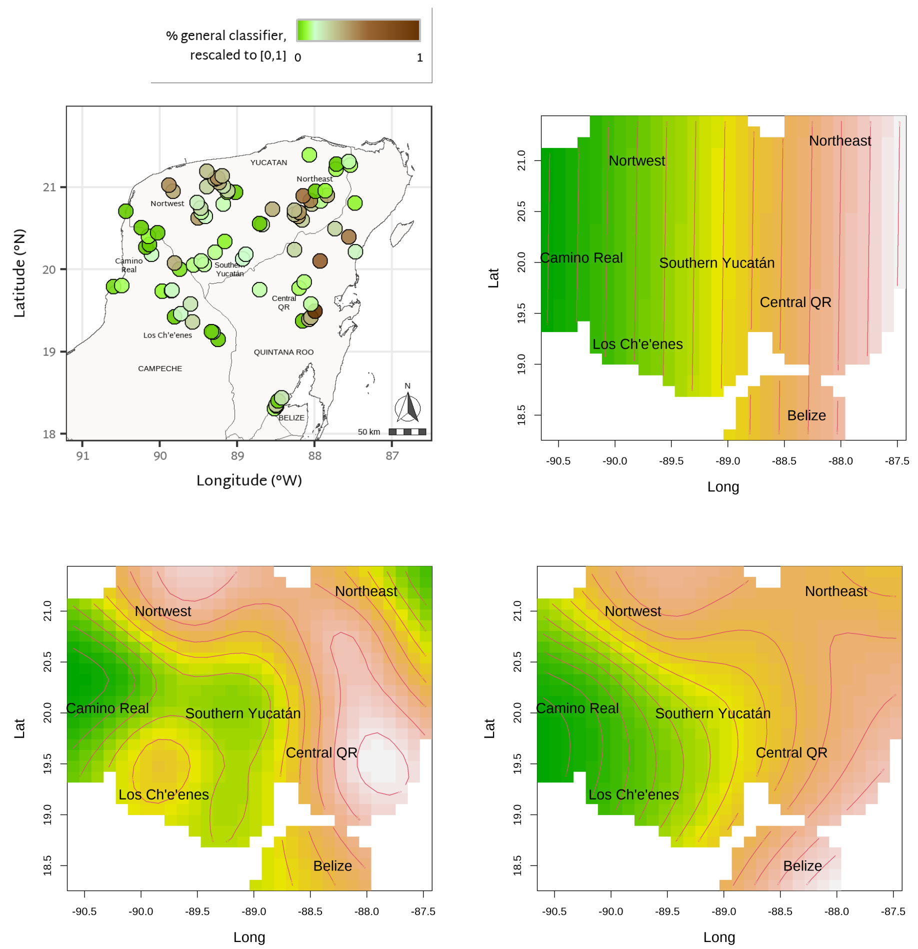

Figure 5 Mensural Classifiers in Space: Top Left Panel: % general classifier (in multiple-classifier constructions) out of total valid responses; Top Right and Bottom Panels: GAM-coefficients of Space on the occurrence of general classifier: *Parts* (Top Right), *Sums* (Bottom Left), *Portions* (Bottom Right).

6 Discussion

The hypotheses in (22) are repeated in (27) for convenience:

(27) Hypotheses about language change

(a) Change in the general classifier
The classifier *p'éel* evolves into a general marker of *Cardinality*.

(b) Change in the mensural classifiers
The mensural classifiers lose their dual nature and are only used as measure nouns.

The analysis of sortal classifiers revealed that:

- The use of the general classifier significantly increases in the West/North.

This difference has been already reported in earlier studies (Martínez Corripio 2003: 85) and can be replicated in corpus data, which do not have the artefacts of the translation task. In narratives collected in the State of Yucatán (Monforte et al. 2010), only 3 out of the first 20 occurrences of the noun *máak* 'person' with numeric expressions are formed with the classifier *p'éel* 'CL.UNIT', while the remaining examples are formed with the classifier *túul* 'CL.ANIM' (7 speakers, age range: 18-84; mean: 44). In narratives collected in Campeche (Can Canul & Gutiérrez Bravo 2016) (Southwest), 10 out of the first 20 numeric constructions with the same noun are formed *p'éel* and the remainder with *túul* (6 speakers; age range: 60-94; mean: 73). It is indicative that the frequency of *p'éel* in this small sample contains a visible effect of age: older speakers (4 speakers; 73-94 years old), use *p'éel* in 3 out of 13 tokens, while younger speakers (3 speakers; 60-65 years old) use *p'eel* in all 7 tokens. Thus, the frequencies in spontaneous data (a) confirm that our findings reflect a genuine dialectal difference between West and East, (b) suggest a recent change in the western varieties (within the 20th century).

The following findings are informative for the relation between the two constructions at issue:

- The probability of substituting sortal classifiers by the general classifier is a significant predictor of the use of the general classifier in mensural constructions (see significant effect of Unit/Sort in Table 4).
- The geographical dispersion differs: substituting sortal classifiers by the general classifier predominates in the West/North, using general

classifiers in mensural constructions predominates in the East/North (compare coefficients of Space in Figure 3 and Figure 5).

The former finding straightforwardly confirms the predictions of (27a). Besides dialectal variation (as captured by the Space coefficients), speakers who frequently substitute sortal classifiers with the general classifier also tend to use the general classifier in mensural classifier constructions. This effect indicates that these phenomena are interconnected and can be accounted for by the view that the classifier *p'éel* evolves into a general marker of *Cardinality*, attaching across-the-board to numeral roots. The latter finding shows that the tendencies in the individual constructions are not in an implicative relation: the use of general classifier increases in the West for sortal classifiers and in the East for mensural classifiers. These findings are not contradictory, as they reflect tendencies at different levels. At the level of individuals, these constructions are correlated, meaning that beyond areal differences, speakers either tend to use general classifiers in both constructions or not. At the level of geographical areas, distinct patterns emerge depending on construction: the general classifier is more established in sortal-classifier constructions in the West and in the mensural-classifier construction in the East.

The following findings in our data confirm the hypothesis in (27b):

- The use of the general classifier in mensural classifier constructions significantly depends on Measure Type (see significant effect of Measure Type in Table 4).
- The dispersion of the general classifier in Space significantly depends on Measure Type (see significant interaction effects between Space and Measure Type in Table 4 and Space coefficients in Figure 5).

The categories *Parts*, *Sums*, and *Portions* are groups of prompts selected in the design of the AYM data in order to test whether the properties of measure expressions generalize across subtypes (see Section 4.2). It is very likely that the differences in our findings are due to lexical differences between individual classifiers (see Figure 4), but the crucial issue is only that such differences exist – even if the exact source of variation cannot be precisely determined by this data. The increase of the frequency of multiple classifiers is part of the process leading from mensural classifiers with a dual nature to measure nouns (loss of the possibility to attach to numeral roots). This phenomenon is already attested with other mensural classifiers in the history of Yucatec Maya, as introduced in Section 2.

The results of both analyses reveal an effect of Spanish Bias:

- The use the general classifier in sortal classifier constructions (Table 2) and mensural classifier constructions (Table 4) significantly depends on the Spanish Bias of the speaker.

Hence, besides dialectal variation and the further examined factors, speakers having a stronger bias of Spanish in this task used the general classifier more frequently in both constructions. This effect reveals a further common property between these constructions, pointing to a possible influence of Spanish. This influence is an instance of pattern borrowing (De León Pasquel 1988: 78): Spanish numerals are expressions of number that can be valued as cardinal without any overt expression of Cardinality; see (9a). The closest expression in Yucatec Maya is a numeral root with the general classifier. As mentioned in Section 2, Mayan numerals with the general classifier slightly differ from Spanish numerals: they can be used with any countable entity, except for kinds.

Finally, the AYM data do not provide evidence for a diachronic trend, since Time, as reflected in the apparent-time scale of the birth years of the speakers was not significant in either analysis. Similarly, there is no evidence for an effect of Population Size, which would be informative for changes between rural and urban contexts (population size is often a predictor of dialectal levelling in urban centers; see discussion in Wieling et al. 2011, Hilton 2010, Blaha Pfeiler & Skopeteas 2022). Hence, we cannot draw inferences about diachronic trends within the time range of the present sample or within the types of centers (villages and small urban centers) that are represented in this study. This does not necessarily imply that such differences do not exist in the population, but that the variation in the *AYM* sample is adequately explained by differences between areas that are the product of earlier diachronic processes.

7 Conclusions

The challenge of dialectal syntax is to infer the historical processes that underlie synchronic variation between dialects. Variation in Space can be the product of change in Time, that is, the dialectal variation that we observe today may be the result of processes that took place earlier than the time range represented by the sample.

Yucatec Mayan numeral roots require numeral classifiers whenever they form cardinal expressions. Two constructions of numeral classifiers are subject to variation in the contemporary language. In sortal-classifier constructions, numeral roots are either accompanied by a specific classifier that identifies the class of the nominal predicate or by a general classifier. In mensural-classifier constructions, numeral roots are either accompanied by a mensural classifier or by a general classifier and a mensural classifier. Based on data collected in 85 locations in the peninsula of Yucatán, we identified a partially overlapping pattern in the distribution of these constructions that can be accounted for by two diachronic processes that are accumulated in contemporary Yucatec Maya.

The first diachronic process is the generalization of the classifier *p'éel*. This element originates in a classifier with limited distribution in Colonial Maya (occurring with coins and non-classified artefacts), which has been desemanticized over time (Bricker 2019: 240–244). As already reported independently of our study, the contemporary language displays two versions of this phenomenon, appearing in different areas: the classifier *p'éel* is used for (all) inanimates in central/eastern varieties and for all entities in the western varieties. The present study shows a stronger preference for substituting specific sortal classifiers by the general one in the West/North – compared to other regions. Furthermore, this preference explains a part of the variation in the use of the general classifier with mensural classifiers. The conjunction of these phenomena leads to a classifier that is applied across-the-board whenever numeral roots are used in cardinal expressions, i.e., a generalized marker of *Cardinality*. The scheme in (28) extrapolates the observed trends (Stage$_2$ is not yet completely reached, since *p'éel* can relate to all atomic entities, but not to kinds). This process is an instance of grammaticalization: the classifier *p'éel* shifts from a classifier restricted to certain classes of nominal predicates to a marker of a functional category. The complement of this process is that the inventory of sortal classifiers shrinks over time, as pointed out in earlier studies (Thompson 1970: 319, Briceño Chel 1992: 71–75, Bricker 2019: 240).

	expression	Cardinality	Sort	Measure
(28)				
	p'éel 'cl.inan' in Stage$_1$	✓	✓	–
	p'éel 'cl.unit' in Stage$_2$	✓	–	–

The second diachronic process lies in the mensural classifiers. These elements have a dual nature in Yucatec Maya: (a) they express the measure to be counted in numeric expressions, which is a property that they share with measure nouns; (b) they can serve as markers of *Cardinality*, directly attaching to numeral roots. The difference between mensural classifiers and measure nouns is that only the former can directly attach to numeral roots; setting this difference apart, mensural classifiers can be used as nouns outside numeral constructions. The effect of the recategorization is that mensural classifiers lose their dual nature and cease to directly attach to numeral roots, as outlined in (29). The present study shows that (a) the likelihood of using an expression of measure with the general classifier inversely correlates with its frequency (general classifiers are more commonly used with rare mensural classifiers); (b) the dispersion of the use of general classifiers in Space partly depends on mensural classifier (effect of Measure Type).

(29)

expression	Cardinality	Sort	Measure
luuch 'cl.gourd' in Stage_1	√	–	√
luuch 'gourd' in Stage_2	–	–	√

Thus, the microvariation in the use of numeral classifiers reveals a rather complex historical scenario. The observed phenomena cannot be reduced to a single source of variation; they can be understood as the result of two diachronic processes that are only partially related. Beyond the purely linguistic factors, variation is determined by the speaker's bias for Spanish and shows up with different patterns in different geographical spaces. The complexity of multi-variate data demonstrates the relevance of examining the dimensions of microvariation in order to test hypotheses about syntactic change. This is not a novel insight, but rather corroborates a recurrent conclusion of microvariationist studies. A basic insight at the outset of dialectological research was that the idealized expectations of the Neo-Grammarians about the regularity of sound change were not confirmed by data from oral non-standardized dialects (Boberg, Nerbonne & Watt 2018: 7). Finally, these findings do not challenge the possibility to obtain generalizations but they are a natural consequence of the increase of the empirical scope: the need for granularity increases along with the size of the data set.

References

Baayen, R. Harald & Maja Linke. 2020. Generalized additive mixed models. In Magali Paquot & Stefan Th. Gries (eds.), *A practical handbook of corpus linguistics*, 563–591. Cham: Springer International Publishing. doi:10.1007/978-3-030-46216-1_23.

Bailey, Guy. 2004. Real and apparent time. In J. K. Chambers, Peter Trudgill & Natalia Schilling-Estes (eds.), *Handbook of language variation and change*, 312–332. Wiley. doi:10.1002/9780470756591.ch12.

Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The apparent time construct. *Language Variation and Change* 3(3). 241–264. doi:10.1017/s0954394500000569.

Bale, Alan & Jessica Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4). 695–707. doi:10.1162/ling_a_00170.

Bale, Alan, Jessica Coon & Nicholás Arcos. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. *Glossa: a journal of general linguistics* 4(1). 77.1–30. doi:10.5334/gjgl.752.

Barrera Vázquez, Alfredo. 2012. Contrata de un maya de Yucatán, escrita en su lengua materna, para servir en Cuba, en 1849. *Estudios de Cultura Maya* 1. 199–209. doi:10.19130/iifl.ecm.1961.1.206.

Barrera Vásquez, Alfredo. 1980. *Diccionario Maya Cordemex*. Merida: Cordemex.

Bastarrachea Manzano, Juan R. & Jorge M. Canto Rosado. 2003. *Diccionario maya popular*. Merida: Academia de la Lengua Maya de Yucatán, A.C.

Statistical Institute of Belize, The. 2010. *Belize population and housing census 2010: Country report*. Belize: Belmopan. www.statisticsbelize.org.bz/aboutsurveys/census2010.

Beltrán, R. P. Fr. Pedro. 1859. *Arte del idioma maya: reducido a sucintas reglas y semilexicon yucateco*. Mérida: J. D. Espinosa. 1st edition 1742.

Benincà, Paola. 1989. Introduction. In Paola Benincà (ed.), *Dialect variation and the theory of grammar*, 1–8. Dordrecht: Foris.

Berlin, Brent. 1968. *Tzeltal numeral classifiers: a study on ethnographical semantics*. The Hague, Paris: Mouton.

Blaha Pfeiler, Barbara. 2021. Yucatec Maya Dialect Atlas. doi:10.5281/ZENODO.5139975. (Dataset).

Blaha Pfeiler, Barbara & Andrew Hofling. 2006. Apuntes sobre la variación dialectal en el maya yucateco. *Península* 1. 27–44.

Blaha Pfeiler, Barbara & Stavros Skopeteas. 2022. Sources of convergence in indigenous languages: lexical variation in Yucatec Maya. *PLOS ONE* 17(5). e0268448. doi:10.1371/journal.pone.0268448.

Blaha Pfeiler, Barbara, Stavros Skopeteas & Elisabeth Verhoeven. 2022. Atlas of Yucatec Maya Online. Online resource. Version 0.0.1. https://spw.uni-goettingen.de/shiny/yuc/.

Boberg, Charles, John Nerbonne & Dominic Watt. 2018. Introduction. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *The Handbook of Dialectology*, 1–18. Wiley Blackwell.

Bochnak, Ryan & Lisa Matthewson. 2020. Techniques in complex semantic fieldwork. *Annual Review of Linguistics* 6. 261–283.

Bohnemeyer, Jürgen. 2015. A practical epistemology for semantic elicitation in the field and elsewhere. In Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in semantic fieldwork*, 13–46. New York: Oxford University Press.

Brandner, Ellen. 2012. Syntactic microvariation. *Language and Linguistics Compass* 6(2). 113–130. doi:10.1002/lnc3.320.

Brandner, Ellen. 2020. A "borderline case" of syntactic variation. *Glossa: a journal of general linguistics* 5(1). doi:10.5334/gjgl.606.

Briceño Chel, Fidencio. 1992. El uso de sufijos clasificadores numerales y de mensurativos en el maya yucateco. *Estudios de Lingüística Aplicada* 15/16,. 69–80.

Briceño Chel, Fidencio. 1993. *La cuantificación en maya: el uso de clasificadores numerales y de mensurativos*. Mérida Universidad Autónoma de Yucatán MA thesis.

Briceño Chel, Fidencio. 1994. La cuantificación actual en el maya yucateco. *Estudios de Lingüística Aplicada* 18/20. 142—150.

Briceño Chel, Fidencio. 2002. Lengua e identidad entre los mayas de la península de Yucatán. *Los Investigadores dela Cultura Maya* 10. 370—379.

Briceño Chel, Fidencio & Gerónimo Ricardo Can Tec. 2014. *U nu'ukbesajil u ts'íibta'al maayat'aan. normas de escritura para la lengua maya*. Mexico, D.F.: INALI.

Bricker, Victoria, Eleuterio Poot Yah & Ofelia Dzul de Poot. 1998. *A dictionary of the Maya language as spoken in Hocabá, Yucatán*. Salt Lake City: The University of Utah Press.

Bricker, Victoria R. 2019. *A historical grammar of the Maya language of Yucatan: 1557-2000*. Salt Lake City: The University of Utah Press.

Bucheli, Claudia & Elvira Glaser. 2002. The syntactic atlas of Swiss German dialects: empirical and methodological problems. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 41–74. Amsterdam: Meertens Institute.

Can Canul, César & Rodrigo Gutiérrez Bravo. 2016. *Maayáaj tsikbalilo'ob Kaampech: Narraciones mayas de Campeche*. Mexico, D.F.: Instituto Nacional de Lenguas Indígenas.

Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6(4). 339–405. doi:10.1023/a:1008324218506.

Cornips, Leonie. 2002. Variation between the infinitival complementizers om/voor in spontaneous speech data compared to elicitation data. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 75–96. Amsterdam: Meertens Institute.

Coronel, Juan. 1620 [1929]. Arte en lengua de maya. : Repr.: (ed.) 1929:. In Juan Martínez Hernández (ed.), *'diccionario de Motul maya español' atribuido a Fray Antonio de Ciudad Real y 'arte de lengua maya' por Fray Juan Coronel*, 2–55. Mérida, Yucatán: Compañia Tipográfica Yucateca.

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45(2). 145–171. doi:10.1080/00437956.1994.11435922.

De León Pasquel, María de Lourdes. 1988. *Nouns and numeral classifiers in*

Mixtec and Tzotzil: a referential view. Brighton: University of Sussex dissertation. https://www.academia.edu/29037961.

Deal, Amy Rose. 2017. Countability distinctions and semantic variation. *Natural Language Semantics* 25(2). 125–171. doi:10.1007/s11050-017-9132-0.

Edmonson, Munro S. 1986. *Heaven born Merida and its Destiny: the book of Chilam Balam of chumayel*. Austin: University of Texas Press. doi:10.7560/719378.

Fishman, Joshua A. 1967. Bilingualism with and without diglossia; diglossia with and without bilingualism. *Journal of Social Issues* 23. 29–38.

Fleischer, Jürg. 2000. Preposition stranding in German dialects. In Sjef Barbiers, Leonie Cornips & Susanne van der Kleij (eds.), *Syntactic microvariation*, 116–151. Amsterdam: Meertens Institute.

García Matzar, Pedro & José Obispo Rodríguez Guaján. 1997. *Rumekir ri kaqchikel chi': gramática kaqchikel*. Guatemala: Cholsamaj.

Grinevald, Colette. 2000. A morphosyntactic typology of classifiers. In Gunter Senft (ed.), *Systems of nominal classification*, 50–92. Cambridge: Cambridge University Press.

Gutiérrez Bravo, Rodrigo & Melanie Uth. 2020. Functional projections and language contact: the case of plural marking of Spanish nouns in Yucatec Maya. *Typology of Morsphosyntactic Parameters* 3. 80–104.

Hilton, Nana Haug. 2010. *Regional dialect levelling and language standards: changes in the Hønefoss dialect.* York: The University of York dissertation.

INEGI. 2011. *Censo de Población y Vivienda 2010: Tabulados del cuestionario básico*. Mexico D.F.: Instituto Nacional de Estadística y Geografía. https://www.inegi.org.mx.

Kayne, Richard S. 1996. Microparametric syntax. In James R. Black & Virginia Motapanyane (eds.), *Microparametric syntax and dialect variation*, ix. John Benjamins Publishing Company. doi:10.1075/cilt.139.01kay.

Keller, Kathryn C. 1955. The Chontal (Mayan) numeral system. *International Journal of American Linguistics* 21(3). 258–275.

Knowles-Berry, Susan M. 1987. Linguistic decay in Chontal Mayan: the speech of semi-speakers. *Anthropological linguistics* 29(4). 332–341.

Krifka, Manfred. 1995. Common nouns: a contrastive analysis of English and Chinese. In Gregory Carlson & Francis Jeffry Pelletier (eds.), *The generic book*, 398–411. Chicago: University of Chicago Press.

Law, Danny & David Stuart. 2017. Classic Mayan: an overview of language in ancient hieroglyphic script. In Judith Aissen, Nora C. England & Roberto Zavala Maldonado (eds.), *The Mayan languages*, 128–172. London: Routledge.

Lehmann, Christian. 1998. *Possession in Yucatec Maya*. Munich: LINCOM. 2nd edition.

Lehmann, Christian. 2010. On the function of numeral classifiers. In Franck Floricic (ed.), *Essais de typologie et de linguistique générale: mélanges offerts à Denis Creissels*, 435–445. Lyon: École Normale Supérieure.

Lehmann, Christian. 2020. Grammaticalization in Yucatec Maya. In Walter Bisang & Andrej Malchukov (eds.), *Grammaticalization scenarios: cross-linguistic variation and universal tendencies, vol. 2: Grammaticalization scenarios from Africa, the Americas, and the Pacific*, 803–852. De Gruyter. doi:10.1515/9783110712735-005.

Lehmann, Christian. 2023. La lengua maya de Yucatán. website. www.christianlehmann.eu/ling/sprachen/maya/index.php.

Little, Carol-Rose, Mary Moroney & Justin Royer. 2022. Classifiers can be for numerals or nouns: Two strategies for numeral modification. *Glossa: a journal of general linguistics* 7(1). 1–35. doi:10.16995/glossa.8437.

Lucy, John A. 1992. *Grammatical categories and cognition: a case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Macri, Martha J. 2000. Numeral classifiers and counted nouns in the Classic Maya inscriptions. *Language and Dialect in the Maya Hieroglyphic Script* 3(1). 13–36. doi:10.1075/wll.3.1.03mac.

Martínez Corripio, Naim Israel. 2003. *Cuatro procesos de cambio lingüístico en el maya yucateco: generalización y pérdida de clasificadores numerales, elisión del clasificador nominal alak, cambio de orden en la frase nominal con adjetivo y cambio de función gramatical en los adverbios*. Mexico DF: Escuela Nacional de Antropología e Historia. [Tesis de Licenciatura].

Mateo Pedro, Pedro. 2022. Adquisición de los clasificadores numerales en q'anjob'al. *Revista Pueblos y fronteras digital* 17. 1–30.

Matthewson, Lisa. 2004. On the methodology of semantic fieldwork. *International Journal of American Linguistics* 70. 369–415.

Miram, Helga-Maria. 1983. *Numeral classifiers im Yukatekischen Maya*. Hannover: Verlag für Ethnologie.

Monforte, Jorge, Lázaro Dzul & Rodrigo Gutiérrez Bravo. 2010. *Narraciones mayas*. Mexico, D.F.: Instituto Nacional de Lenguas Indígenas.

Moser, Ann-Marie. 2019. Form und Funktion der doppelten Negation in deutschen Dialekten, mit einem Schwerpunkt im Oberdeutschen. *Linguistik Online* 98(5). 179–195. doi:10.13092/lo.98.5935.

Pfeiler, Barbara. 2009. The acquisition of numeral classifiers and optional plural marking in Yucatec Maya. In Ursula Stephany & Maria D. Voeikova (eds.), *Development of nominal inflection in first language acquisition: a cross-linguistic perspective*, 91–110. Mouton de Gruyter. doi:10.1515/9783110217117.91.

Pfeiler, Barbara. 2014. Maya and Spanish in Yucatán: an example of continuity and change. In Salikoko S. Mufwene (ed.), *Iberian imperialism and language evolution in Latin America*, 205–224. Chicago: The University of Chicago Press.

Poletto, Cecilia. 2000. *The higher functional field: evidence from Northern Italian dialects*. New York, Oxford: Oxford University Press.

Polian, Gilles. 2013. *Gramática de tseltal del Oxchuc, tomo II*. Mexico City: Centro de Investigaciones y Estudios Superiores en Antropología Social.

Pool Balam, Lorena & Olivier Le Guen. 2015. La integración de préstamos en maya yucateco: Una perspectiva interaccional acompañada de una reflexión sobre la ideología del lenguaje. *Amerindia* 37. 343–384.

R Core Team. 2023. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Scontras, Gregory. 2013. Accounting for counting: A unified semantics for measure terms and classifiers. In *Proceedings of SALT 23*, 549–569.

Scontras, Gregory. 2022. On the semantics of number morphology. *Linguistics and Philosophy* 45(5). 1165–1196. doi:10.1007/s10988-021-09345-8.

Skopeteas, Stavros, Amedee Colli Colli, Daniela Schellenbach, Carolin Brokmann, Florian Fischer & Maya Gálvez Wimmelmann. 2020. Yucatec Maya spoken corpus. The Language Archive, Corpus resource. https://hdl.handle.net/1839/00-0000-0000-0021-E91B-F.

Thompson, J. Eric S. 1970. Sufijos numerales y medidas en yucateco. *Estudios de Cultura Maya* 8. 319–339. doi:10.19130/iifl.ecm.1972.8.341.

Tonhauser, Judith & Lisa Matthewson. 2016. Empirical evidence in research on meaning. The Ohio State University and University of British Columbia.

Tozzer, Alfred. 1921. *A Maya Grammar*. Cambridge, MA: Peabody Museum.

Uth, Melanie & Rodrigo Gutiérrez Bravo. 2018. La marcación de número en los préstamos españoles del maya yucateco: variación y restricciones. *Cuadernos de Lingüística de El Colegio de México* 5(1). 183–224.

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2022. itsadug: interpreting time series and autocorrelated data using GAMMs. R package version 2.4.1.

Vázquez-Rojas Maldonado, Violeta, Josefina García Fajardo, Rodrigo Gutiérrez-Bravo & Julia Pozas Loyo. 2018. The definite article in Yucatec Maya: the case of le ...o'. *International Journal of American Linguistics* 84. 207–242.

Whalen, Gretchen. 2003. *An annotated translation of a Colonial Yucatec manuscript: on religious and cosmological topics by a native author*. FAMSI.

http://www.famsi.org/reports/01017/01017Whalen01.pdf.
Wieling, Martijn, Simonetta Montemagni, John Nerbonne & R. Harald Baayen. 2014. Lexical differences between Tuscan dialects and standard Italian: accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language* 90(3). 669–692.
Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: explaining linguistic variation geographically and socially. *PLoS ONE* 6(9). e23613. doi:10.1371/journal.pone.0023613.
Wieling, Martijn, Esteve Valls, R. Harald Baayen & John Nerbonne. 2018. Border effects among Catalan dialects. In Dirk Speelman, Kris Heylen & Dirk Geeraerts (eds.), *Mixed effects regression models in linguistics* Quantitative Methods in the Humanities and Social Sciences, 71–97. Springer. doi:10.1007/978-3-319-69830-4_5.
Wolk, Christoph & Benedikt Szmrecsanyi. 2018. Probabilistic corpus-based dialectometry. *Journal of Linguistic Geography* 6(1). 56–75. doi:10.1017/jlg.2018.6.
Wood, Simon N. 2010. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73(1). 3–36. doi:10.1111/j.1467-9868.2010.00749.x.
Wood, Simon N. 2017. *Generalized Additive Models*. Chapman and Hall/CRC. doi:10.1201/9781315370279.
Yu, Yidong. 2023. Optionality, variation and categorial properties: The case of plural marking in Yucatec Maya. In Kristin Kopf & Thilo Weber (eds.), *Free variation in grammar*, 284–314. John Benjamins Publishing Company. doi:10.1075/slcs.234.10yu.
Yu, Yidong. 2024. *Pluralizing entities and events in Yucatec Maya*. Göttingen: University of Göttingen dissertation. doi:http://dx.doi.org/10.53846/goediss-10468.
Zabbal, Youri. 2005. The syntax of numeric expressions. Ms, UMass. Amherst.
Zavala, Roberto. 2000. Multiple classifier systems in Acatec (Mayan). In Gunter Senft (ed.), *Systems of nominal classification*, 114–146. Cambridge University Press.
Zuur, Alain F., Elena N. Ieno, Neil J. Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.

Supplementary material

- *BlahaSkopeteas2024Supplementary.pdf*: report generated from the R Markdown file.
- *YUC-CL.rmd*: R Markdown file containing the code and analysis workflow.
- *YUC-CL.rda*: Processed dataset used in the analysis, including all annotations.

Abbreviations

Ø	zero	DEP	dependent status
3	3rd person	DIST	distal
A	person marker A	IMP	imperative
ADJR	adjectivalizer	INAN	inanimate
ANIM	animate	INCMPL	incompletive status suffix
B	person marker B	IPFV	imperfective
CL	class	LOC	locative
CMPL	completive	PFV	perfective
COLL	collective	PL	plural
CONJ	conjunction	POSS	possessor
CONT	continuator	REL	relationalizer
DEF	definite	TOP	topic
		TRR	transitivizer

Appendix I: Prompts

Prompt labels: question identifiers in the *Atlas of Yucatec Maya* (Blaha Pfeiler 2021).

A. Prompts for Sorts (5 prompts)

Q111 *llevaron un hombre a una cueva* 'they took one/a man to one/a cave'
Q163 *una cochino* 'one/a pig'
Q164 *una piedra* 'one/a stone'
Q165 *una mata de huaya* 'one/a huaya plant'
Q166 *una vela* 'one/a candle'

B. Prompts for Portions (*5 prompts*)

Q167 *una bola de masa* 'one/a ball of dough'
Q170 *un puño de cemento/fertilizante/maiz* 'one/a fistful of cement/fertilizer/corn'
Q178 *un montón de excremento* 'one/a pile of excrement'
Q180 *dos tragos de agua* 'two shots of water'
Q187 *una gota de agua* 'one/a drop of water'

C. Prompts for Sums (*7 prompts*)

Q169 *una fila de piedras* 'one/a row of stones'
Q174 *un montón, estiba de papel/tortilla/ropa* 'one/a pile, stack of paper/tortillas/clothes'
Q179 *una carga de leña* 'one/a load of firewood'
Q181 *un manojo de hierbas* 'one/a bunch of herbs'
Q182 *una doblada de ropas* 'one/a folded bundle of clothes'
Q184 *dos pesadas de frijol* 'two portions of beans'
Q272 *un amarre de frijol* 'one/a mooring of beans'

D. Prompts for Parts (*7 prompts*)

Q171 *una mitad de sandía* 'one/a half of watermelon'
Q172 *un pedazo de piedra/vidrio/plástico* 'one/a piece of stone/glass/plastic'
Q173 *un pedazo quebrado de madera* 'one/a broken piece of wood'
Q177 *un pedazo rasgado de ropa* 'one/a torn piece of clothing'
Q183 *dos rebanadas de queso* 'two slices of cheese'
Q185 *un pedazo de ropa* 'one/a piece of clothing'
Q186 *un pedazo cortado de madera* 'one/a cut piece of wood'

Appendix II: Classifiers

Glosses are based on the available studies and dictionaries of Yucatec Maya (Bricker et al. 1998, Bastarrachea Manzano & Canto Rosado 2003, Barrera Vásquez 1980, Beltrán 1859, Miram 1983, Briceño Chel 1992). The symbol "(*)" means that the classifier is not reported in these sources; the gloss was inferred from its use in the AYM data.

A. Classifiers obtained in prompts for Sorts

classifier	*n*	GLOSS
che'	1	TREE
kúul	60	PLANT
p'éel	261	UNIT
ts'íit	43	LONG
túul	200	ANIMATE
xéek	3	FOOT OF TREE

classifier	*n*	GLOSS
TOTAL	568	

B. Classifiers obtained in prompts for Portions

classifier	*n* without *p'éel*	*n* with *p'éel*	sum	% with *p'éel*	GLOSS
ch'áaj	88	14	102	13,7	DROP
ch'ooj	15	4	19	21,1	TWIST
cháach	2	1	3	33,3	HANDFUL
chúuch	1	0	1	0,0	AGGREGATE (*)
chuuk'	1	0	1	0,0	SPOONFUL
jeneb	41	1	42	2,4	PILE
k'ab	2	0	2	0,0	HAND
kóots	1	0	1	0,0	ROLLED PIECE
láap'	28	6	36	17,6	FISTFUL
lóob	0	1	1	100,0	FISTFUL (*)
lóoch'	7	4	11	36,4	SCOOP
lóot	3	0	3	0,0	HANDFUL
luuch	2	0	2	0,0	HALF SQUASH
luuk'	87	23	110	20,9	MOUTHFUL
maach	1	0	1	0,0	HANDFUL
múuch'	10	0	10	0,0	HEAP
múul	3	0	3	0,0	PILE
nikib	3	0	3	0,0	PILE
p'u'uk	6	0	6	0,0	VESSEL
p'úuy	0	1	1	100,0	FRAGMENT
t'aaj	6	0	6	0,0	BIT
ts'úuk	1	1	2	50,0	LUMP
tukub	7	2	9	22,2	PILE
wóoch'	1	0	1	0,0	CLUSTER
wóok	8	1	9	11,1	CLUSTER
wóol	74	20	94	21,3	BALL
xa'ak	0	1	1	100,0	BASKET
xuuch	12	3	15	20,0	VESSEL
TOTAL	410	83	493	16,8	

C. Classifiers obtained in prompts for Sums

classifier	*n* without *p'éel*	*n* with *p'éel*	sum	% with *p'éel*	GLOSS
beel	3	0	3	0,0	WAY
ch'uuy	2	0	2	0,0	HANGING POT
cháach	82	18	100	18,8	HANDFUL
chooj	1	0	1	0,0	CLUSTER

classifier	*n* without *p'éel*	*n* with *p'éel*	sum	% with *p'éel*	GLOSS
chúuch	3	2	5	40,0	AGGREGATE (*)
chúuj	0	1	1	100,0	GOURDFUL
jaats	0	1	1	100,0	DIVISION
jiil	3	1	4	25,0	STRAND
juuts'	2	0	2	0,0	FOLD (*)
k'áax	80	16	96	16,7	TIED BUNDLE
k'ab	1	1	2	50,0	HAND
kúuch	114	27	141	19,1	LOAD
láap'	9	4	13	30,8	FISTFUL
lóoch'	2	1	3	33,3	SCOOP
lóot	3	1	4	25,0	HANDFUL
maach	18	1	19	5,3	HANDFUL
meek'	1	0	1	0,0	ARMLOAD
múuch'	3	3	6	50,0	HEAP
múul	1	1	2	50,0	PILE
muut	3	0	3	0,0	MEASURE OF CORN
p'iis	42	13	55	23,6	LOAD
p'óoch	0	2	2	100,0	BUNCH
paak	34	12	46	26,1	FOLDER
t'i'in	0	1	1	100,0	STRING
t'o'ol	8	5	13	38,5	LINE
t'úul	0	1	1	100,0	ROW
to'	1	0	1	0,0	BUNDLE
ts'áam	1	0	1	0,0	PAIR
ts'ap	62	23	85	27,1	STACK
tsóol	54	26	80	32,5	LINE
tukub	3	0	3	0,0	PILE
waats'	1	1	2	50,0	TIMES
wóoch'	1	0	1	0,0	CLUSTER
wóok	3	0	3	0,0	CLUSTER
wóol	6	1	7	14,3	BALL
wuuts'	10	4	14	28,6	FOLD
TOTAL	557	167	724	23,1	

D. Classifiers obtained in prompts for Parts

classifier	*n* without *p'éel*	*n* with *p'éel*	sum	% with *p'éel*	GLOSS
búuj	75	0	75	0,0	HALF
ch'áak	1	0	1	0,0	CUT PIECE
cháach	1	0	1	0,0	HANDFUL
jáat	47	1	48	2,1	PIECE
káach	49	4	53	7,5	SLIVER
kóots	10	0	10	0,0	ROLLED PIECE

classifier	*n* without *p'éel*	*n* with *p'éel*	sum	% with *p'éel*	GLOSS
p'aay	10	0	10	0,0	CRUMB
p'úuy	1	0	1	0,0	FRAGMENT
t'i'il	1	0	1	0,0	THREAD
táaj	3	0	3	0,0	PIECE
táan	5	2	7	28,6	HALF
tséej	4	0	4	0,0	SLIVER
tsíil	1	0	1	0,0	SHRED
wáat	1	0	1	0,0	PART
wóol	0	1	1	100,0	BALL
xéek	1	0	1	0,0	PLANT/FOOT
xéet'	503	23	526	4,4	PIECE
xóot'	106	2	108	1,9	CUT PIECE
TOTAL	819	33	852	3,9	

Barbara Blaha Pfeiler
Universidad Nacional Autónoma de México, CEPHCIS
Calle 43 s/n entre 44 y 46, col. Industrial
C.P. 97150 Mérida, Yucatán, México
blahadegler@gmail.com

Stavros Skopeteas
Universität Göttingen
Käte-Hamburger-Weg 3
37073 Göttingen, Germany
stavros.skopeteas@uni-goettingen.de